ATATÜRK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANA BİLİM DALI

Serhat BURMAOĞLU

BİRLEŞMİŞ MİLLETLER KALKINMA PROGRAMI BEŞERİ KALKINMA ENDEKSİ VERİLERİNİ KULLANARAK DİSKRİMİNANT ANALİZİ, LOJİSTİK REGRESYON ANALİZİ VE YAPAY SİNİR AĞLARININ SINIFLANDIRMA BAŞARILARININ DEĞERLENDİRİLMESİ

DOKTORA

TEZ YÖNETİCİSİ

Prof.Dr.Erkan OKTAY

ERZURUM-2009

TEZ KABUL TUTANAĞI

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bu çalışma, İşletme Anabilim Dalının Sayısal Yöntemler Bilim Dalında jürimiz rafından Doktora tezi olarak kabul edilmiştir.

Prof.Dr. Erkan OKTAY
Danışman/Jüri Üyesi

Prof.Dr. Vedat PAZARLIOĞLU
Jüri Üyesi

Doç.Dr. Uğur YAVUZ
Jüri Üyesi

Doç.Dr. Hüseyin ÖZER
Jüri Üyesi

Y.Doç.Dr. M.Suphi ÖZÇOMAK
Jüri Üyesi

Yukarıdaki imzalar adı geçen öğretim üyelerine aittir. 11 / 12 / 2009

Prof.Dr. Mustafa YILDIRIM
Enstitü Müdürü

İÇİNDEKİLER

ÖZET

DOKTORA TEZİ

BİRLEŞMİŞ MİLLETLER KALKINMA PROGRAMI BEŞERİ KALKINMA ENDEKSİ VERİLERİNİ KULLANARAK DİSKRİMİNANT ANALİZİ, LOJİSTİK REGRESYON ANALİZİ VE YAPAY SİNİR AĞLARININ SINIFLANDIRMA BAŞARILARININ DEĞERLENDİRİLMESİ

Serhat BURMAOĞLU

Danışman: Prof.Dr.Erkan OKTAY

2009- SAYFA: 116

Jüri : Prof.Dr.Erkan OKTAY

Prof.Dr.Vedat PAZARLIOĞLU

Doç.Dr.Hüseyin ÖZER

Doç.Dr.Uğur YAVUZ

Y.Doç.Dr.M.Suphi ÖZÇOMAK

Günümüzde yapılan araştırmalar incelendiğinde artık tek değişkene bağlı olarak olayların analizi yerine birçok farklı değişkenin dikkate alındığı görülmektedir. Bu tez çalışmasında çok değişkenli istatistik yöntemleri ve yapay sinir ağları kullanılarak sınıflandırma analiz yöntemlerinin sınıflandırma gücü test edilmiştir. Analizde Birleşmiş Milletler Kalkınma Programı tarafından Beşeri Kalkınma Endeksinin hesaplanmasında kullanılan değişkenlerden yararlanılmıştır. Yapılan üç analizin sonucunda Çok Katmanlı Perseptron Modeli ve Lojistik Regresyon Modelinden Diskriminant Analizine göre daha iyi sınıflandırma başarısı elde edilmiş, Diskriminant Analizinden ise tatminkâr bir sonuç elde edilmiştir. Diskriminant Analizinde normallik varsayımları için logaritmik dönüşümler kullanılmış ve bu sayede yaklaşık 1 puanlık sınıflandırma başarısı artırılmıştır. Çok Katmanlı Perseptron Modelinde ise ham veriler kullanıldığında elde edilen sınıflandırma oranının geliştirilmesi için değişkenlere ait değerler normalize edilmek suretiyle %100'lük sınıflandırma başarısı elde edilmiştir.

ABSTRACT

Ph.D. THESIS

EVALUATING CLASSIFICATION SUCCESS OF DISCRIMINANT ANALYSIS, LOGISTIC REGRESSION ANALYSIS AND NEURAL NETWORK MODELS USING UNITED NATIONS DEVELOPING PROGRAMME'S HUMAN DEVELOPMENT INDEX

Serhat BURMAOĞLU

Supervisor : Prof.Dr.Erkan OKTAY

2009- PAGE: 116

Jury : Prof.Dr.Erkan OKTAY

Prof.Dr.Vedat PAZARLIOĞLU

Assoc.Prof. Hüseyin ÖZER

Assoc.Prof. Uğur YAVUZ

Asist. Prof. M.Suphi ÖZÇOMAK

When the studies investigated nowadays, on behalf of analyzing cases up to the one variable, it can be seen that more than one and different variables taken into consideration. Classification power of classification analysis methods has been tested in this thesis study by using multivariate statistical techniques and neural network models. Human Development Index data, used by United Nations Developing Programme (UNDP) for classifying countries, have been used in the analysis. After making the analysis, better classification has been made by Multi Layer Perceptron Model and Logistic Regression Model. Also Discriminant Analysis produces a satisfactory result too. Classification success of Discriminant Analysis method has increased 1 point by providing uni-variate normality with logarithmic transformations of some variables. For improving classification success of Multi Layer Perceptron Model, variables are normalized. Because of normalization of variables, 100 percent classification success has been reached. Finally, results are discussed and which factors have affected the methods is scrutinized.

ŞEKİLLER DİZİNİ

Sayfa No

ÇİZELGELER DİZİNİ

Sayfa No

GİRİŞ

Günlük hayatta karşılaşılan olaylar genellikle tek değişkene bağlı olmadan birçok değişkenle açıklanabilecek karmaşıklıktadır. Bu sebeple yapılan araştırmalarda tek değişkenli istatistik tekniklerin sınırlamalarından kaçınmak maksadıyla çok değişkenli istatistiksel analizlerin kullanımına başlanmıştır.

Çok değişkenli istatistik metotlar inceleme konusu olayı bir bütün olarak ele almakta ve bütünlüğü sağlayan değişkenlerin bağımlılık yapısını açıklamaya çalışmaktadır. Bu durumda çok değişkenli istatistik metotların en önemli amacının değişkenler arasındaki bağımlılığın yapısının analizi olduğu iddia edilebilir (Tatlıdil 1996). Çok değişkenli analizde birden çok değişkenin analizi ile ilgilenildiğinden en az iki değişken söz konusudur. Çok değişkenli analiz ile ilgili olarak literatürde belirtilen şu tanımlamalar dikkat çekicidir:

- Çok değişkenli analiz, eşzamanlı olarak çok sayıda bağımlı değişken veya bağımlı ve bağımsız değişken ayrımı yapılmadan çok sayıda değişkenle ilgilenildiğinde söz konusu olmaktadır (Shin 1996).
- Çok değişkenli analiz, örnek üzerinde ikiden fazla değişkeni eşzamanlı çözümleyen tüm istatistik tekniklerdir (Sheth 1971).
- Çok değişkenli analiz, değişken grupları arasındaki karşılıklı ilişkileri ölçme ve açıklama olanağı veren istatistik tekniklerdir (Gatty 1966; Albayrak 2006).
- Çok değişkenli analiz teknikleri, her birim için çok sayıda birbirinden farklı, ancak birbiriyle ilişkili değişkenlerin ölçüldüğü ve bütün değişkenlerin eşzamanlı olarak dikkate alınıp birlikte incelendiği tekniklerdir (Albayrak 2006).

İfade edilen tanımların ortak özellikleri incelendiğinde ikiden fazla olmak kaydıyla çok sayıda değişken ve bu değişkenleri ölçerken kullanılacak bir bağımlı değişkenin varlığı gözlenmektedir. Kısacası değişken grupları arasındaki karmaşık ilişkilerin ortaya çıkarılmasında çok değişkenli istatistik analiz yöntemlerinin kullanıldığı ve çok değişkenli analizin amacının da bu ilişkisel boyutun farklı perspektiflerden incelenmesi amacına dönük olduğu söylenebilir.

Çok değişkenli analiz teknikleri farklı ölçütlere göre sınıflandırılabilmektedir. Analiz teknikleri; karşılıklı bağımlılık veya bağımlılık analizi yapıp yapmadığına, çıkarsama veya doğrulama gibi iki farklı amaca dönük olup olmadığına ve son olarak kullandığı verinin tipine göre sınıflandırılmaktadır.

Sharma (1996: 6) veri analiz metotlarını bağımlı yöntemler ve karşılıklı bağlılık yöntemleri olarak ayrıma tabi tutmuştur. Bağımlı yöntemler Çizelge 1'de gösterilmektedir. Ölçülebilir ve ölçülemeyen değişkenlerin bir veya birden fazla olmasına göre kullanılacak teknikler gruplandırılmıştır.

Çizelge 1 Bağımlı Yöntemler (Sharma 1996)

Bağımsız Değişkenler	**Bağımlı Değişkenler**			
	Bir Değişken		**Birden Fazla Değişken**	
	Ölçülebilir	**Ölçülemeyen**	**Ölçülebilir**	**Ölçülemeyen**
Bir Değişken Ölçülebilir	• Regresyon	• Diskriminant Analizi • Lojistik Regresyon	• Kanonik Korelasyon	• Çok Gruplu Diskriminant Analizi (ÇGDA)
Ölçülemeyen	• t-test	• Kesikli Diskriminant Analizi	• Çok Değişkenli Varyans Analizi	• Kesikli ÇGDA
Birden Fazla Değişken Ölçülebilir	• Çoklu Regresyon	• Diskriminant Analizi • Lojistik Regresyon	• Kanonik Korelasyon	• ÇGDA
Ölçülemeyen	• Varyans Analizi	• Kesikli Diskriminant Analizi • Bitişiklik Analizi	• Çok Değişkenli Varyans Analizi	• Kesikli ÇGDA

Çizelge 1'de belirtilen bağımlı yöntemlere karşın karşılıklı bağlılık yöntemlerinde ise iki değişken sayısı için ölçülebilir verilerde basit korelasyon kullanılırken, ölçülemeyen veri türlerinde loglineer modeller kullanılmaktadır. İkiden daha fazla değişkenlerde ise ölçülebilir verilerde Temel Bileşenler Analizi ve Faktör Analizi kullanılmakta, ölçülemeyen veri türlerinde yine Loglineer Modeller ve Karşılık Getirme Analizleri kullanılmaktadır.

Tez çalışmasında çok değişkenli istatistiksel analiz tekniklerinden iki tanesi kullanılacağından çok değişkenli istatistik yöntemlerinden burada kısaca bahsedilerek

genel çerçeve çizilmeye çalışılacaktır. Ayrıca teknikler arasındaki ilişkilerde burada kısaca izah edilecektir.

Temel Bileşenler Analizi, çok değişkenli verinin boyutunun azaltılması için kullanılan bir yöntemdir. Boyut indirgeme yoluyla tek başına bir analiz yöntemi olarak kullanılabildiği gibi başka analizler için veri hazırlama tekniği olarak ta kullanılabilir. Temel bileşenler analizi ile çok boyutlu ve açıklanması karmaşık olan bir araştırmanın daha az boyuta indirgenerek açıklanabilir hale getirilmesi hedeflenmektedir.

Faktör analizinde de yine temel bileşenler analizine benzer bir şekilde boyut indirgeme yolu ile araştırmanın açıklanabilir hale getirilmesi hedeflenmektedir. Temel bileşenler analizinden farkı ise araştırmada incelenen çok sayıdaki değişkenin birbirleri ile olan ilişkilerinin incelenerek ortaklıklar kurulması ve tek bir faktör altında bunların toplanabilmesidir. Amaç değişkenler arasındaki bağımlılığın kökenini araştırmaktır. Örneklem büyüklüğü faktör analizi için önemlidir. Gözlem sayısı değişken sayısından fazla olmalıdır (Akgül 1997). Faktör analizi değişkenler arasındaki korelâsyonlardan yola çıkarak değişkenleri sınıflandırmaya tabi tutar. Bu sayede aynı sınıf içerisindeki değişkenler arasında yüksek korelasyon bulunurken farklı sınıflar arasındaki değişkenler arasında düşük korelasyon bulunacaktır. Bu sayede birbirleriyle ilişki içerisinde olan değişkenlerin ortak özelliklerinden yola çıkılarak veri açıklanabilir hale getirilmiş olur.

Kümeleme analizi, kategorize etme ile ilgilenir. Gözlem yapılan büyük grubu göreceli olarak benzer nitelikte küçük gruplara bölerek bu gruplardan işe yarar bilgiler elde edilmesini amaçlamaktadır. Kümeleme analizi yapılma süreci ile ilgili olarak birçok metot kullanılmaktadır. Kümeleme analizinin faktör analizinden farkı kümeleme analizinde gözlemlenen değişkenlerin kati suretle ortak bir özelliğe sahip olanlara göre kümelenmiş olmasıdır. Faktör analizinde değişkenler arasındaki korelasyona göre kümeleme yapılmaktadır.

Çok Değişkenli Varyans Analizi (ÇDVA), deneysel faktör veya faktörlerin bağımlı değişkenler üzerindeki etkisini test etmeyi hedefleyen bir tekniktir. Ölçümlerin çok değişkenli normal dağılıma uygun olması ve varyansların homojen olması arzu edilir. Bu sayede kurulacak hipotez ile farklılık yaratan gruplar bulunabilecektir.

Diskriminant Analizi diskriminant fonksiyonu kullanılarak gruplar arası ayırıma en fazla etki eden değişkenleri belirlemede ve hangi gruptan geldiği bilinmeyen bir bireyin hangi gruba dâhil edileceğini belirlemede kullanılır. Genel anlamda ayırma olup,

bireylere ait *p* tane özellikten yararlanarak ait oldukları grupları (kütle) belirlemede veya mevcut grupları birbirinden ayıracak en iyi fonksiyonu bulmada kullanılan çok değişkenli istatistik tekniklerden birisidir (Çamdeviren 2000).

Diskriminant analizi bağımlı değişkenin kesikli olması durumunda bağımlılık analizinin ölçülmesinde daha iyi bir yöntem olarak gözükmektedir. Diskriminant analizinin sonucu bağımlı değişkenin farklı değerleri arasında ayrım yapmaya imkân veren doğrusal bir fonksiyondur.

Lojistik Regresyon Analizi bir veya birden fazla bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi modelleyerek gözlemleri lojistik ayırt etme fonksiyonunu kullanarak gruplara atamaktır. Tek değişken varsa lojistik regresyon çok değişken varsa çoklu regresyon söz konusu olur. Diskriminant analizinin normallik varsayımının geçerli olmadığı durumlarda lojistik regresyon kullanılır.

Çok Boyutlu Ölçekleme Analizi, veri kümesi içerisindeki nesneler arasında göreceli bir yakınlık veya benzerlik olması durumunda kullanılır. Bu veriden yararlanarak tüm araştırmadaki verilerin ilişkisel haritasını çıkararak genel resmin gösterilmesi hedeflenmektedir. Bu analiz belli bir dağılım varsayımı gerektirmeyen bir yöntemdir. Kümeleme analizine benzer matematiksel alt yapısı bulunmaktadır.

Karşılık Getirme Analizi (Conjoint Analysis), doğrulama ve test maksadıyla kullanılan bir yöntemdir. Özellikle kategorik verilerin analizinde kullanılır. Örnek olarak 20X20'lik bir matris olduğu düşünüldüğünde, bu büyüklükte bir tablonun yorumlanması zor olacaktır. Ancak matrisin satır ve sütunlarını özetleyebilecek birkaç bileşenin bulunması işleri kolaylaştıracaktır. Karşılık getirme analizi bu amacı gerçekleştirmektedir. Temel bileşenler analizine benzemekte hatta ölçülemeyen değişkenlerle yapılan temel bileşenler analizi olarak ta karşılık getirme analizi ifade edilmektedir.

Kanonik Korelâsyon Analizi, çok değişkenli iki küme arasındaki ilişkileri inceleyen bir yöntemdir. Temel bileşenler analizinde bir kümedeki çok değişken dikkate alınırken kanonik korelâsyonda iki kümedeki değişkenler dikkate alınmaktadır. Ancak temel bileşenler analizinde olduğu gibi kanonik korelâsyon analizinde de hedef, boyutun indirgenerek konunun açıklanabilir hale getirilmesidir.

Belirtilen izahatlardan da anlaşılabileceği gibi bazı teknikler karmaşık değişken yapısını benzerliklerden yararlanmak kaydıyla indirgeyerek kolay açıklanabilir hale

getirmeyi amaçlamakta, bazıları da yine benzerlikleri kullanarak sınıflandırma ve doğru tahmin yapmak amacıyla kullanılmaktadır. Sınıflandırma yöntemleri de incelendiğinde iki grubun oluştuğu görülmektedir. Sınıfların önceden bilinen gruplar olması veya önceden grupların bilinmemesi durumuna göre sınıflandırma teknikleri de kendi içlerinde ikiye ayrılmaktadır. Sınıfların önceden bilinmemesi durumuna göre sınıflandırmada çok boyutlu ölçekleme analizi ve kümeleme analizi kullanılırken, sınıfların önceden bilinmesi durumunda ise diskriminant analizi ve lojistik regresyon analizi kullanılmaktadır.

Bu çalışmada önceden bilinen ve Birleşmiş Milletler Kalkınma Programı tarafından yapılan çok gelişmiş ve orta düzeyde gelişmiş ülke sınıflandırması dikkate alındığından yeniden sınıflandırma için diskriminant analizi ve lojistik regresyon analizi yöntemleri kullanılmıştır. Ayrıca yapay sinir ağları da kullanılarak sınıflandırma başarısı incelenmiştir. Sınıflandırmada kullanıldığı belirtilen üç teknik bir, iki ve üçüncü bölümlerde daha detaylı olarak ele alınmış, matematiksel alt yapıları, varsayımları ve işlem mantığı izah edilmeye çalışılmıştır.

Dördüncü bölümde uygulamada kullanılan veri seti ve bu veri setine ilişkin bilgiler verilmiş, bilahare analizler yapılarak sonuçları yorumlanmıştır. Analizde Birleşmiş Milletler Kalkınma Programının, ülkeleri çok gelişmiş ve orta düzeyde gelişmiş sınıflandırmasında kullandığı verilerden yararlanılmıştır. Analizler ise SPSS 16.0 paket programı yardımıyla yapılmıştır.

Sonuçta ikili (binary, dichotomous) bağımlı değişken ile yapılan analizler neticesinde yapay sinir ağlarında öğrenme örnekleminin yüksek tutulması ve verilerin normalize edilmesi sonucu %100'lük bir sınıflandırma başarısı elde edildiği, lojistik regresyon analizinde veriler normalize edilmemesine rağmen yine %100'lük bir başarı elde edildiği, diskriminant analizinde ise verilerin normalize edilerek normallik varsayımı karşılanmasına rağmen %92,5'lik sınıflandırma başarısı elde edildiği görülmüştür. Diskriminant analizinin varsayımlarının fazla olması analiz gücünü ve analiz kolaylığını güçleştirmektedir. Ancak elde edilen sınıflandırma başarı oranları, kullanılan tüm sınıflandırma yöntemlerinin sınıflandırmada önemli bir şekilde yüksek performansa sahip olduklarını göstermiştir.

Bu çalışmanın amacı diskriminant analizi, lojistik regresyon analizi ve yapay sinir ağlarının sınıflandırma başarılarının incelenmesidir. Ancak yapılan bu tez çalışması ile şu konularda da literatüre katkı sağlandığı düşünülmektedir:

- Diskriminant analizi ile yapılan çalışmalar (tez, makale) incelendiğinde bir çok araştırmacının varsayımların bazılarını test etmeksizin doğrudan analizi yaparak sonuçlarını yorumlamaya çalıştığı gözlenmiştir. Bu çalışma ile diskriminant analizi uygulamasında örnek bir yöntem detaylı bir şekilde gösterilmiş,
- Ülkelerin İnsani Kalkınmışlık Endeksinin hesaplanmasında ayırt edici değişkenlerin ne olduğu ve daha az değişken ile bir diğer ifade ile daha az maliyet ile yüksek başarıya sahip sınıflandırmanın hangi modeller kullanılarak yapılacağı belirtilmiş,
- Lojistik regresyon ve yapay sinir ağları modellerinin değişkenlerle ilgili varsayımları olmadığından metrik veriler analizde kullanıldığında daha iyi sonuçlar verdiğine işaret edilmiş,
- Ülkelerin gelişmiş ülke sınıfına dâhil olabilmeleri için hangi konulara ağırlık vermeleri gerektiği belirtilmiştir.

BİRİNCİ BÖLÜM

1. DİSKRİMİNANT ANALİZİ

Çok değişkenli analizde sıklıkla karşılaşılan problemlerden birisi sınıflandırma sorunudur. Araştırmacı farklı yığınlardan gelen bireylerin p sayıdaki özelliğini ölçtüğünde elindeki bireyin hangi gruptan geldiğini merak edebilir. Bu durumda sınıflandırma problemi, bireyin p sayıda özelliğini inceleyerek hangi gruptan geldiğine karar verme problemi olarak nitelendirilebilir.

Sınıflandırma problemi istatistiksel bir karar verme sürecidir. Bu süreçte araştırmacı, bireyin hangi gruptan geldiğine karar vermelidir. Bazı durumlarda grupların olasılık dağılımları ve bu dağılımların parametreleri bilinmektedir. Ancak uygulamada genellikle her grubun p değişkene ilişkin bir dağılıma sahip olduğu varsayılır ve bu dağılımın parametreleri seçilen örnek aracılığıyla tahmin edilir. Ardından karar verme problemi çözülmeye çalışılır. Bu seviyede, araştırmacı için iki karar verme konusu bulunmaktadır. Birincisi grubun ayırt edici özelliklerini araştırarak ayırt edicilikte etkili olan değişkenleri belirlemek ve ikincisi bireyleri bu ayırt edici fonksiyonlar yardımıyla gruplara sınıflandırmaktır.

İstatistiksel bir yöntem olarak diskriminant analizi ilk kez 1936 yılında Ronald A. Fisher tarafından tanıtılmıştır (Albayrak 2006). Bu tarihten itibaren özellikle sınıflandırma ve diğer birçok istatistik araştırmalarda kullanılmaktadır.

Diskriminant analizi ile amaç, hatalı sınıflandırma olasılığını en aza indirgeyerek birimleri ait oldukları gruplara atamak ve gelmiş oldukları ana kütleleri belirlemektir.

Diskriminant analizi X veri setindeki değişkenlerin iki veya daha fazla gerçek gruplara ayrılmasını belirlemek amacıyla yararlanılan bir yöntemdir. Birimlerin p tane karakteristiği ele alınarak bu birimlerin doğal ortamdaki gerçek gruplarına, sınıflarına en uygun düzeyde atanmalarını sağlayacak fonksiyonlar bulunulmasına çalışılır. Bunu yaparken hatalı sınıflandırma maliyetlerini en aza indirgeyerek bireyleri ait oldukları gruplara ayırmak, çekilmiş oldukları kitleleri belirlemek amaçlanır.

Diskriminant analizi, tek faktör çok değişkenli varyans analizinin uzantısı olan çok değişkenli bir analiz türüdür. Gruplar arası fark yoktur anlamını taşıyan H_0 hipotezi reddedildikten sonra, gruplar arası farkın olduğu sonucuna varılır. Bu farklılığın ana nedenleri diskriminant analizi tekniğiyle ortaya çıkarılır (Tabachnick ve Fidell 2001).

Diskriminant analizi aracılığıyla elde edilen diskriminant (ayırıcı) fonksiyonları, bağımsız değişkenlerin doğrusal bileşenlerinden oluşur. Diskriminant fonksiyonları gruplar arası farklılığa etki eden tahmin değişkenlerinin hangileri olduğunu ortaya çıkarır. Gruplar arası farklılığa etki eden bu değişkenlere de diskriminant (ayırıcı) değişkenler adı verilir. Diskriminant analizinin bir diğer işlevi ise, gruplardan herhangi birisine ait olan fakat hangi gruptan geldiği bilinmeyen bir birimin ait olduğu grubu en az hata ile saptamaktır.

Diskriminant analizi, farklılığın en fazla hangi değişkenlerde yoğunlaştığının belirlenmesi ve böylece grupların farklılaşmasında etkin olan faktörlerin saptanmasını da sağlar. Analiz sonucunda yapılan sınıflama ile orijinal grup üyeliklerinin karşılaştırılması, bilinen fonksiyonun yeterli olup olmadığını test etmeye olanak sağlar (Erçetin 1993).

Diskriminant analizi, bir araştırmacının aynı anda çeşitli değişkenlere göre iki veya daha fazla örnek grubu arasındaki farklılıkları çalışmasına olanak sağlayan bir istatistiksel tekniktir. Genel olarak birimlerin gruplanmasında bazı matematiksel eşitliklerden faydalanılır. Diskriminant fonksiyonu olarak adlandırılan bu eşitlikler birbirine en çok benzeyen grupları belirlemeye olanak sağlayacak şekilde grupların ortak özelliklerini belirlemek amacıyla kullanılmaktadır. Grupları ayırmak amacıyla kullanılan karakteristikler ise diskriminant değişkenleri olarak adlandırılmaktadır. Kısaca, diskriminant analizi, iki veya daha fazla sayıdaki grubun farklılıklarının diskriminant değişkenleri vasıtasıyla ortaya konması işlemidir. Birbiriyle yakından ilişkili birkaç istatistiksel yaklaşımı kapsayan geniş bir kavramdır (Klecka 1980).

Diskriminant analizi grupların birbirlerinden en iyi şekilde ayrımını sağlar. Diskriminant analizini kullananlar, açıklama ve tahminle ilgilenirler. Bu gruplarla en fazla ilişkili olan değişkenlerin hangileri olduğunu ve bunların grup üyeliğini ne kadar iyi tahmin edebildiğini belirlemek isterler (Akgül ve Çevik 2003).

1.1. Diskriminant Analizi ve Çoklu Diskriminant Analizinin Amaçları

Diskriminant analizi ve çoklu diskriminant analizinin çeşitli amaçları vardır. Bunlar:

- Örnek durumları bir diskriminant tahmin denklemi ile gruplara ayırmak,

- Örnek durumların tahmin edildiği gibi sınıflandırılıp sınıflandırılmadığına bakarak teoriyi test etmek,
- Gruplar arasında ve gruplar içindeki farkları incelemek,
- Gruplar arasında ayrım yapabilmek için en kolay yolu belirlemek,
- Bağımlı değişkendeki değişimin yüzdesini bağımsız değişkenler yardımıyla belirlemek, (Diskriminant analizi bağımlı değişkenin iki seçenekli bir rastsal değişken olduğunu varsayar.)
- Adımsal diskriminant analizini kullanarak kontrol değişkenleri tarafından hesaplanan bağımsız değişkenler yardımıyla bağımlı değişkendeki varyansın yüzdesini belirlemek,
- Bağımlı değişkeni sınıflandırmada bağımsız değişkenlerin birbirine göre önemini değerlendirmek,
- Grupları ayırt etmede fazla işe yaramayan değişkenleri elemek.

Albayrak (2006) diskriminant analizinin amaçlarını şu şekilde belirtmektedir:

- Ayırıcı değişken setine göre önceden belirlenmiş iki veya daha çok grubun ortalamalarının anlamlı farklılık gösterip göstermediğini belirlemek,
- İki veya daha fazla sayıdaki grubu birbirinden ayıran en önemli değişkenleri saptamak,
- Ayırıcı değişkenlere göre birimleri diskriminant değerlerine göre sınıflandırmak için prosedürler geliştirmek,
- Bağımsız değişkenler tarafından şekillendirilen gruplar arasındaki ayrım boyutlarının kompozisyonunu ve sayısını belirlemek,
- Sınıflandırma uygunluğunun değerlendirmesini yapmak.

İlk üç amacı gerçekleştirmek için yapılan diskriminant analizine, grupları tanımlama amaçlı diskriminant analizi, dördüncü amacı gerçekleştirmek üzere yapılan diskriminant analizine ise karar verme amaçlı diskriminant analizi denilmektedir (Albayrak 2006).

Araştırmada en önemli safha araştırma yapılan konu ile ilgili çıkan sonuçları kullanarak yorum yapmaktır. Diskriminant analizi ile en kuvvetli ayırt edicileri belirleyerek sınıflandırmalar yapmak için matematiksel modeller türetilmektedir. Bu denklemler diskriminant fonksiyonu adını alır. Bu fonksiyonun bulunmasında

belirlenecek grupların ortalaması arasındaki farklılığın maksimum olması amaçlanmaktadır.

Gruplar arasındaki ayrımı gerçekleştirmede kullanılan özellikler ayırt edici değişkenler adını alan bağımsız değişkenlerdir. Bu değişkenler, ölçümün belli bir aralığında veya oranı seviyesinde ölçülmelidir. Bu da ortalama ve varyansların hesaplanabileceği ve matematiksel denklemlerde kullanılabileceği anlamına gelir. Gruplara ayrılacak durum sayısı ile değişken sayısı arasındaki fark ikiden çok olmadıkça ayırt edici değişkenlerin sayısı ile ilgili bir kısıt yoktur.

1.2. Diskriminant Analizinin Varsayımları

Diskriminant analizi, çok değişkenli varyans analizi yönteminde olduğu gibi grupları ortalamalarına (ortalama vektörlerine) göre ortak ortalamadan (ortalama vektöründen) farklı olmalarını sağlayacak bir ayırma kriteri geliştirmeyi amaçlayan bir yöntemdir. Bu nedenle veri setlerine Diskriminant analizi uygulanabilmesi için veri setlerinin aşağıdaki varsayımları taşıması gereklidir.

- X veri matrisi çok değişkenli normal dağılım göstermelidir.
- Değişkenlerin varyans ve kovaryansları homojen olmalıdır. X matrisinde yer alan değişkenler ortak kovaryans matrisine sahip çok değişkenli ana kütleden çekilmiş örnekler olmalıdır.
- Değişkenler arasında çoklu bağlantı (multicollinearity) bulunmamalıdır.
- X matrisi grupların birbirinden ayrılmasında rol oynamayacak gereksiz değişken içermemeli, grupların birbirinden ayrılmasını sağlayacak kadar doğru ve gerekli değişkenleri içermelidir.

Belirtilen varsayımlar genel olarak yapılacak olan istatistiksel analiz için ihtiyaç duyulacak matematiksel modelin ortaya konabilmesi açısından önem taşımaktadır. Aksi takdirde elde edilecek sonuçlar gerçekleri yansıtamayacaktır.

Klecka (1980) diskriminant analizi ile ilgili varsayımları şu şekilde belirlemektedir:

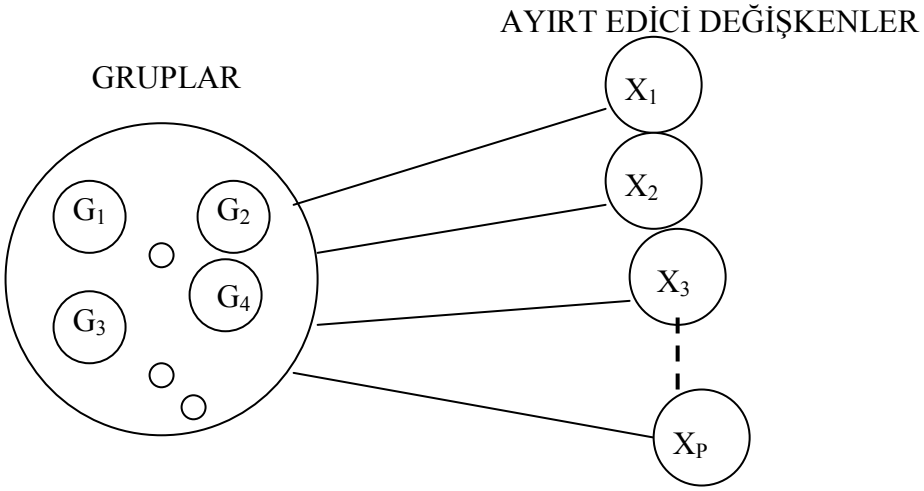
- Hiçbir değişken diğer ayırt edici değişkenlerin doğrusal birleşimi şeklinde bulunmayacaktır.

- Birbiriyle kusursuz ilişki içerisinde olan iki değişken aynı zamanda kullanılmamalıdır.
- Kitle kovaryans matrisleri her grup için eşit olmalıdır.
- Yığından alınan tüm gruplar çok değişkenli normal dağılıma uygun olmalıdır.

Bu sayede grup üyeliği olasılığı ve testin önem derecesi kusursuz bir şekilde hesaplanabilecektir.

Sonuçta diskriminant analizine başlamadan önce normallik varsayımı, kovaryans matrislerinin eşitliği varsayımı ve çoklu bağlantı varsayımı test edilmeli ve elde edilen sonuca göre analize başlanmalıdır.

Anlaşıldığı üzere diskriminant analizi iki veya daha fazla gruplar ve ayırt edici değişkenler arasında farklar üzerinde çalışmaktadır. Bu ilişki Şekil 1.1'de görülmektedir.



Şekil 1.1. Gruplar ve Ayırt Edici Değişkenler Arasındaki İlişki

Görülebileceği gibi gruplar arasında sebebe bağlı, yönlü bir ilişki kullanılmamaktadır. Gruplar bağımlı veya bağımsız değişkenler olarak belirlenmemektedir. Ayrıca ayırt edici değişkenlerde bağımlı veya bağımsız değişken olarak gösterilmemektedir. Şayet araştırma durumu grup kategorilerini ayırt edici değişkenlere bağımlı olarak gösterirse bu durumda çoklu regresyon kullanılabilecektir. Aralarındaki önemli fark diskriminant analizinin nominal düzeyde ölçekle ölçülmüş olan bağımlı değişkeni yönetmesidir.

Diskriminant analizinin varsayımları matematiksel notasyon kullanılarak incelendiğinde;

g= grup sayısı

p= ayırt edici değişken sayısı

n_i= i grubundaki olay sayısı

n= tüm gruplardaki olay sayısı olmak üzere varsayımlar;

- İki veya daha fazla grup $g \geq 2$
- Her grup için en az iki olay $n_i \geq 2$
- Veriler ana kütleden rastgele seçilmiştir.
- Toplam durum sayısının 2 eksiğinden az olmamak kaydıyla herhangi bir sayıda ayırt edici değişken olmalıdır.
- Ayırt edici değişkenler nominal ölçek seviyesinde ölçülür.
- Hiçbir ayırt edici değişken diğer ayırt edici değişkenin doğrusal birleşimi olamaz.
- Özel formüller kullanılmadıkça her bir grup için kovaryans matrisleri eşittir.
- Gruplara ait ortalamalar ve kovaryans matrisi önceden bilinir. Grupların kovaryans (sapma) matrisleri eşittir. Bu varsayımın sağlanmadığı durumlarda diskriminant analizinin kuadratik formülü kullanılabilir.
- Her bir grup, ayırt edici değişkenler üzerindeki çok değişkenli normal dağılıma sahip bir örnek grubundan alınır. Bağımsız değişkenler çok boyutlu normal dağılıma sahiptirler.
- Grupların eşit sayıda birimden oluşmadığı durumlarda, üyelerin tahmini olasılıklarının bilindiği varsayılır.
- Herhangi bir durumun yanlış sınıflandırılmasının maliyeti önceden bellidir.

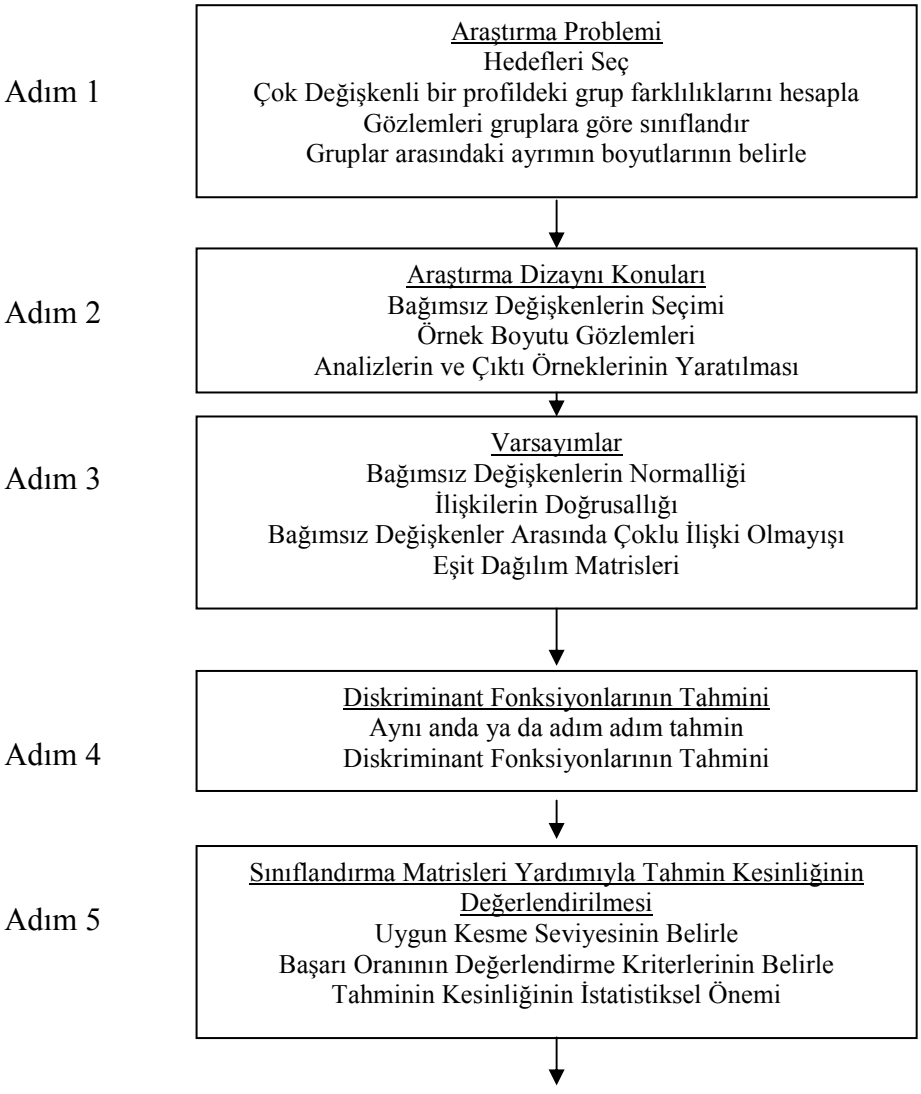
Diskriminant analizi, örnek büyüklüğünün bağımsız değişkenlere oranına (n/p) karşı oldukça duyarlıdır. Birçok çalışma her bir açıklayıcı değişken için 20 birim oranını önermektedir. Pratikte bu oranı sağlamak zor olsa da araştırmacı açıklayıcı değişkenlere göre birim sayısı azaldıkça sonuçların duyarlılığını kaybedeceğini bilmelidir. Minimum örnek büyüklüğü olarak ise her bir açıklayıcı değişken için 5 birim önerilmektedir. Bu kriterler analizdeki nihai değişkenlere göre verilmektedir. Örneğin, bu oran adımsal diskriminant analizinde diskriminant fonksiyonundaki açıklayıcı değişken sayısını göstermektedir (Albayrak 2006). Tez çalışmasında yapılan uygulamada bu oran 1/9'dur. Yani 1 açıklayıcı değişken için 9 birim kullanılmıştır.

Diskriminant analizinde genel örnek büyüklüğünün yanında, her grup için örnek büyüklüğünün de dikkate alınması gerekmektedir. En küçük grup, en az bağımsız

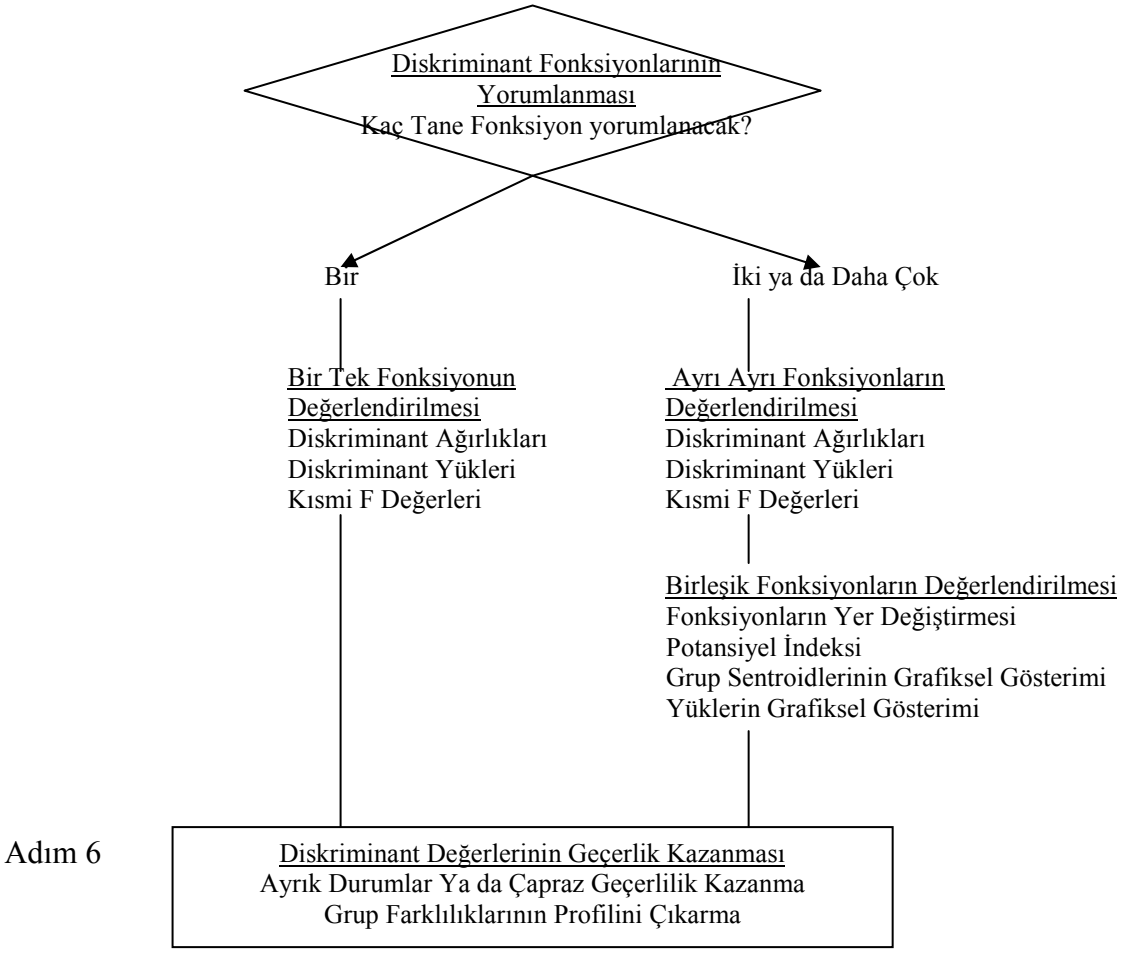
değişken sayısından fazla birim içermelidir. Grup örnek sayıları anormal şekilde farklılık gösterirse, grupların örnek büyüklüğünü diğer gruplara yaklaştırmak amacıyla büyük gruplardan tesadüfî örnekler seçilebilir.

1.3. Diskriminant Analizi İçin Karar Süreci

Diskriminant Analizi için karar süreci Cangül (2006) tarafından 6 adımda toplanmıştır. Şekil 1.2'de bu süreç şematik olarak gösterilmiştir.



Şekil 1.2. Diskriminant Analizi İçin Karar Süreci



Şekil 1.2. (Devam) Diskriminant Analizi İçin Karar Süreci

Diskriminant Analizinin uygulama adımlarını Akgül ve Çevik (2003) ise şu şekilde belirlemiştir:

- Önsel grup üyelikleri belirlenir.
- Analize alınan gruplar arasında fark olup olmadığı Bartlett'in Ki-kare test istatistiği ile elde edilir. Test sonucunda gruplar arasında anlamlı bir fark varsa analize devam edilir.
- Kullanılacak değişkenler seçilir. Değişken seçimine önsel bilgi ya da istatistikî yöntemler uygulanabilir.
- Değişkenler arasında çoklu bağlantı olup olmadığı incelenir. Bu matristeki korelasyon matrisi incelenir. Bu matristeki korelasyon değerleri mutlak değer olarak %75'ten büyükse değişkenlerden bir kısmının atılması gerekir. Bu adımın sonunda değişken kümesi belirlenmiş olur.

- $W^{-1}B$ (*W:* Grup içi kareler toplamı; *B:* Gruplar arası kareler toplamı) matrisinin özdeğerleri ve bu özdeğerlere ilişkin özvektörler bulunur. Bu özvektörler, diskriminant fonksiyonları için gerekli ağırlıkları verir. Diskriminant fonksiyonlarının anlamlılık testi de bu özdeğerler kullanılarak yapılır. Eğer herhangi bir fonksiyon anlamlı ise yapılan ayırımın başarılı olduğu söylenebilir.
- Standartlaştırılmamış diskriminant fonksiyonu kullanılarak her bir birey için diskriminant fonksiyonu değeri elde edilir. Bu değerler sınıflandırma aşamasında kullanılacaktır.
- Grup üyelikleri için önsel olasılıklar belirlenir. Daha sonra, diskriminant fonksiyonunun başarısı doğru sınıflandırma yüzdesi incelenerek tespit edilebilir.

1.4. Diskriminant Analizinde Uygulanan Testler

Diskriminant analizinin belirtilen varsayımlarının karşılanma durumunun kontrolü amacıyla varsayımların sağlanıp sağlanmadığı bir takım testlerle araştırılmalıdır. Bu testler;

- Çok Değişkenli Normallik Testi,
- Grup Kovaryans Matrislerinin Eşitliği Testi,
- Grup Ortalama Vektörlerinin Farklılığı Testi,
- Diskriminant Fonksiyonunun Anlamlılığı Testi,
- Diskriminant Fonksiyonunun Dışsal Geçerlilik Testleridir.

Bu kısımda belirtilen testler sırasıyla ve detaylı olarak incelenecektir.

1.4.1. Çok değişkenli normallik testi

Çok değişkenli normallik varsayımının geçerli olup olmadığının kontrol edilmesindeki amaç, gözlemlerin çok değişkenli normal ana kütlelerden gelip gelmediklerinin test edilmesi ve bunun sonucunda gözlem verileri üzerinde uygun düzeltmelerin yapılabilmesidir. Normallik testlerinde tüm değişkenlerin tek değişkenli normallik testleri yapılarak normal dağılıma uygunluğu kontrol edilir.

Tek değişkenli normalliğin test edilmesinde iki farklı yöntem bulunmaktadır. Bu yöntemlerden birisi grafik testlerdir. Grafik testlerde histogram, gövde-yaprak, kutu,

Q-Q ve P-P grafikleri kullanılmaktadır. Bir diğer yöntem ise analitik testlerin kullanılmasıdır. Analitik testlerden sıklıkla kullanılanlar ise Ki-kare uygunluk, Kolmogorov-Smirnov (K-S) ve Shapiro-Wilks testleridir. Tezin uygulama bölümünde Kolmogorov-Smirnov testi tek değişkenli normalliğin test edilmesinde kullanılmıştır. Kolmogorov-Smirnov testi ile

H_0: Seçilen Örneklem Normal Dağılıma Uygundur

H_A: Seçilen Örneklem Normal Dağılıma Uygun Değildir.

hipotezi test edilmektedir. Hipotezden de anlaşılabileceği gibi başlangıç hipotezinin kabul edilmesine çalışılır. Test sonucunda normal olarak kabul edilen değişkenler analiz için kullanılacak, normal dağılım göstermeyen değişkenler ise öncelikle çarpıklık ve basıklık açısından incelenerek dönüşüme tabi tutulacaktır.

Normal dağılım göstermeyen değişkenler logaritmik, karekök veya *ln* dönüşümleri ile normalleştirilmeye çalışılır. Çünkü çok değişkenli normal dağılım gösteren bir veri setinin tüm değişkenleri tek değişkenli normal dağılıma sahiptir. Ancak bunun tersinin geçerli olmadığı Hair vd (1995) tarafından ifade edilmiştir.

Şayet ***x****, p* elemanlı rastlantı değişken vektörü; μ ortalama vektörü, Σ kovaryans matrisi ile çok değişkenli normal dağılıma sahipse $\mathbf{x} \sim N_p(\mu;\Sigma)$ biçiminde gösterilir.

$$f(x) = ke^{-\frac{1}{2}(\frac{x-\mu}{\sigma})} = \frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad (1.1)$$

biçiminde yazılan olasılık yoğunluk fonksiyonu çok değişkenli olması halinde;

$$f(x) = k\exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]; -\infty < \mu_j \in \mu < \infty; -\infty < x_j \in x < \infty \qquad (1.2)$$

$$\Sigma > 0;\ j = 1,\ldots\ldots,\ p \text{ için}$$

olarak gösterilecektir. Burada *pxp* boyutlu, simetrik, pozitif Σ matrisinin elemanları ile *px1* boyutlu μ ortalama vektörünün elemanları sonludur ve dağılımın parametreleridir. Fonksiyondaki *k* sabiti, *x* vektör değişkeninin tanımlandığı bölgede integrali 1 yapan değer olacaktır.

Bir veri setinin çok değişkenli normalliğinin test edilmesi için D^2 uzaklık ölçüsü (Mahalanobis uzaklık ölçüsü) kullanılır.

$$D^2_{ik} = (X_i - X_k)' S_w(x)^{-1} (X_i - X_k) \qquad (1.3)$$

Burada S ortak gruplar içi kovaryans matrisi olup,

$S_w=(1/(N-m))W$'dir. (1.4)

Bu yöntemin en büyük avantajı *p* tane değişkene doğrudan uygulanabilmesidir. Uygulama aşamaları ise sırasıyla şu şekildedir:

- D^2_{ij}, uzaklık kareleri hesaplanır,
- Bulunan değerler büyükten küçüğe doğru sıralanır,
- D^2_{ij}, $p(i-1/2)$n serbestlik derecesiyle X^2dağılımı gösterir ve D^2_{ij} çiftlerinin grafiği çizilir, bu doğru ki-kare grafiği olarak adlandırılır.
- Elde edilen ki-kare grafiğinin şeklinin bir doğru oluşturması halinde çok değişkenli normalliğin geçerliliğine karar verilir (Sharma 1996: 381).

Ayrıca çok değişkenli normal dağılımın analizinde bir diğer yöntem de sapan değerlerinin incelenmesidir. Çok değişkenli sapan birimler, analizde kullanılacak bağımsız değişkenler arasındaki kareli Mahalanobis uzaklıkları (MD^2) hesaplanarak saptanabilmektedir. Şöyle ki; analizdeki her birim için hesaplanan kareli Mahalanobis uzaklıkları analizdeki değişken sayısına (p, serbestlik derecesine) bölünerek elde edilen değer (MD^2/df) t dağılımına uymaktadır. Herhangi bir birimin sapan değer olarak değerlendirilebilmesi için ilgili birimin %1 anlamlılık düzeyinde anlamlı olması gerekmektedir. Yani MD^2/df değerinin (t) 5,014'den büyük olması gerekmektedir (Hair vd 1995: 66-70). Sapan değer analizi aslında bir çeşit uç değer analizi olarak ta değerlendirilebilir.

1.4.2. Grup kovaryans matrislerinin eşitliği testi

Grupların kovaryans matrislerinin eşit olması durumunda "doğrusal diskriminant fonksiyonları", farklı olması durumunda ise "kuadratik diskriminant fonksiyonları" uygun olmaktadır. Bu nedenle diskriminant analizine başlamadan önce grup kovaryans matrislerinin eşitliği test edilmektedir. m tane p değişkenli normal grubun kovaryans matrislerinin eşitliğini test etmek için

H_0:$\Sigma_1 = \Sigma_2 = \ldots\ldots = \Sigma_m$ hipotezi ele alınsın. Bu hipotezi test etmek için k gruptan N_k, $k=1,2,\ldots..,m$ sayıda bireyden oluşan bir rastgele örnek için S_k örnek kovaryans matrisi, Σ_k'nin sapmasız bir tahmini olsun. H_0 hipotezi doğru ise ortak kovaryans matrisi,

$S=(N-m)\sum_{k=1}^{m}(N_k-1)S_k \quad ; \quad N=\sum_{k=1}^{m}N_k$ şeklinde olacaktır. Bu durumda test istatistiği şu şekilde ifade edilir:

$$V=(N-m)\ln|S|-\sum_{k=1}^{m}(N_k-1)\ln|S_k| \qquad (1.5)$$

ve,

$$C^{-1}=1-\frac{2p^2+3p-1}{6(p+1)(m-1)}\left|\frac{1}{\sum(N_k-1)}-\frac{1}{(N-m)}\right| \qquad (1.6)$$

Ölçü faktörü kullanılarak VC^{1} miktarının örnek hacmi büyüdükçe $1/2(m-1)p(p+1)$ serbestlik dereceli bir ki-kare dağılımına yaklaştığı gösterilmiştir.

$$VC^{-1}\le\chi_{1/2(m-1)p(p+1)} \qquad (1.7)$$

Bu durumda H_0 hipotezinin kabul edileceği yani kovaryans matrislerinin farklı olmadığı kararına varılır. SPSS uygulamalarında kovaryans matrislerinin eşitliği Box M testi ile

H_0: Kovaryans Matrisleri Eşittir.

H_A: Kovaryans Matrisleri Birbirinden Farklıdır.

hipotezi test edilerek yapılmaktadır. Box M Kovaryans Matrislerinin Eşitliği/ Homojenliği testinde F dağılımını kullanmaktadır. Box M testinde kullanılan test istatistiği ise,

$$M=\left\{\prod_{i=1}^{n}|C_i|^{(n_i-1)/2}\right\}/|C_i|^{(n-m)/2} \qquad (1.8)$$

dir. Yazılan eşitlikte C_i i'inci örneklemin kovaryansını, *n* toplam gözlem sayısını, n_i ise i'inci örneklemin büyüklüğünü göstermektedir. Test sonucunda başlangıç hipotezinin kabul edilmesine çalışılır.

Ayrıca kovaryans matrislerinin homojenliğinin test edilmesinde Bartlett testinden de yararlanıldığı bilinmektedir. Bartlett test istatistiği *p* değişken sayısı ve *g* grup sayısı olmak üzere *p(p+1)(g-1)/2* serbestlik derecesinde χ^2 tablo değeri ile karşılaştırılarak test edilmektedir.

1.4.3. Grup ortalama vektörlerinin farklılığı testi

Grup ortalama vektörlerinin farklılığı da bir diğer diskriminant analizi varsayımıdır. Grup ortalama vektörleri arasında önemli bir fark olmaması diskriminant analizini gereksiz kılabilecektir. Çünkü bu durumda grupları tam anlamıyla birbirinden ayırmak mümkün olamayacak ve bireylerin gruplara sınıflanmasında yeterince iyi ölçüler geliştirilemeyecektir.

Bu amaçla,

$H_0 : \mu_1 = \mu_2 = ... = \mu_m$ hipotezine karşılık

H_A: En az iki μ birbirinden farklıdır hipotezi test edilir.

H_0 hipotezini test etmek için, S.S.Wilks tarafından geliştirilen ve Wilks'in Lambdası olarak adlandırılan test istatistiği kullanılır. Bu istatistik,

$$\Lambda = \left|\frac{W}{T}\right| \text{'dir.} \qquad (1.9)$$

M.S.Bartlett Λ istatistiğine dayanarak p(m-1) serbestlik derecesiyle yaklaşık ki-kare istatistiğini şu şekilde geliştirmiştir:

$\chi^2_{p(m-1)} = -\left[N - 1 - (p+m)/2\right] \ln \Lambda$ bu eşitlik sonrasında hesaplanan χ^2 değeri, α=0,05 anlamlılık düzeyinde p(m-1) serbestlik derecesine karşılık gelen χ^2 tablo değerinden büyükse, $\chi^2_H > \chi^2_{p(m-1)}$ en az iki grup ortalama vektörünün birbirinden farklı olduğuna karar verilecektir.

1.4.4. Diskriminant fonksiyonunun anlamlılığı testi

Diskriminant fonksiyonunun istatistiksel olarak anlamlılığının test edilmesi de gruplar arası ayrımın daha az boyutlu betimlenmek istendiğinden önemlidir. Bunun için gruplar arasındaki farkın önem kontrolünde kullanılan Wilks test istatistiğinden yararlanılır. (Λ) test ölçütü, λ ayırma ölçütüne bağlı olarak aşağıdaki gibi bulunur:

$$1/\Lambda = |T|/|W| = |W^{-1}T| \qquad (1.10)$$

T yerine $W+B$ konursa,

$$1/\Lambda = |W^{-1}(W+B)| \qquad (1.11)$$

$$= \left| I + W^{-1} B \right|$$

$$= \prod_{i=1}^{r} (1 + \lambda_i)$$

Buradan, $\Lambda = 1 / \prod_{i=1}^{r} (1 + \lambda_i)$ ilişkisine varılır. Bartlett, (Λ) ölçütünü kullanarak, p(m-1) serbestlik derecesiyle yaklaşık ki-kare dağılımını gösteren aşağıdaki test istatistiğini geliştirmiştir:

$$\chi^2_{p(m-1)} = \left[N - 1 - (p + m)/2 \right] \ln \Lambda \qquad (1.12)$$

$$= \left[N - 1 - (p + m)/2 \right] \sum_{i=1}^{r} \ln(1 + \lambda_i)$$

Bu istatistik seçilen α anlamlılık düzeyinde $\chi^2_{p(m-1)}$ tablo değeri ile karşılaştırılarak, gruplar arasındaki farkın önemli olup olmadığı test edilir. Gruplar arasındaki farklılık istatistiksel açıdan önemli ise diskriminant fonksiyonlarından en az birisinin önemli olduğuna karar verilir (Ünal 2006).

1.4.5. Diskriminant fonksiyonunun dışsal geçerlilik testi

Diskriminant fonksiyonunun dışsal geçerlilik testleri incelendiğinde Alıkoyma (Holdout) Metodu, U-metodu ve Tekrarlı Örnekleme (Bootstrap) yöntemi kullanıldığı görülmektedir (Sharma 1996).

Alıkoyma metodunda örnek tesadüfî olarak iki örneğe bölünmektedir. Diskriminant fonksiyonu bu iki örnekten birisiyle hesaplanıp diğer örneğe ait birimler sınıflandırılarak fonksiyonun dışsal geçerliliği test edilmektedir. Elde edilen sonucun sınıflama hatasının tarafsız tahminleyeni olması gerekmektedir. Bu yöntem çift çapraz geçerlilik testi olarak ta bilinmektedir. Bu yöntemin uygulanabilmesi için örneğin büyük olması gerekmektedir.

U-metodu, Lachenbruch tarafından 1967 yılında ileri sürülmüş, her adımda bir gözlem değerinin örnekten çıkarılıp geriye kalan n-1 gözlem değeriyle diskriminant fonksiyonu tahmin edilip dışarıdaki gözlem değerlerinin sınıflandırılmasına dayanmaktadır. Bu n tane diskriminant analizinin çözülmesi ve dışarıda tutulan n tane gözlem değerinin sınıflandırılması demektir. Yöntem, hata kareleri ortalaması (veya varyansı) minimum olan bir hata oranı sağlamadığı için eleştirilmiştir (Albayrak 2006).

Tekrarlı örnekleme yöntemi, bir örnekten tekrarlı örnekler alınmasına dayanan bir yaklaşımdır. Diskriminant analizi bu tekrarlanan örnekler üzerine uygulanarak bir hata oranı hesaplanmaktadır. Genel hata oranı ve örnekleme dağılımı tekrarlanan örneklerin hata oranlarından elde edilmektedir.

1.5. Parametrik Diskriminant Analizi Teknikleri

Parametrik diskriminant analizi teknikleri çok değişkenli normal dağılım ve grup kovaryans matrislerinin benzerliği gibi iki önemli koşulu varsayım olarak ele almak suretiyle analizde kullanan tekniklerdir. Dolayısıyla çok az ya da hemen hemen hiçbir varsayım gerektirmeyen olaylarda kullanılamaz. *p* sayıda değişkene göre kitleleri ya da kitle içindeki alt grupları belirlemek için çoğu zaman parametrik çok değişkenli diskriminant analiz tekniklerinden Doğrusal Diskriminant Analizi (Linear Discriminant Analysis) veya Kuadratik Diskriminant Analizi (Quadratic Discriminant Analysis)'nden yararlanılmıştır (Öztürk 2006).

Bazı araştırmacılar diskriminant analizinde ayırma fonksiyonu katsayılarının hesaplanmasında başvurulan yöntemlere göre diskriminant analizi isminin başına getirilen ek sözcüklere göre Fisher'in Doğrusal Diskriminant Analizi, Kernel Tabanlı Kümeleme ile Diskriminant Analizi(Kernel Based Discriminant Analysis), En Büyük Benzerlik Diskriminant Analizi (Maximum Likelihood Discriminant Analysis), Bayes Diskriminant Analizi (Bayesian Discriminant Analysis), Laplacian Doğrusal Diskriminant Analizi (Laplacian Linear Diskriminant Analysis) gibi isimlerle anmayı uygun bulmaktadırlar (Tang vd 2005; Liang ve Shi 2004; Lu vd 2005; Zheng 2005; Srivastava vd 2007).

Doğrusal diskriminant analizi teknikleri veri sınıflandırma ve boyut indirgeme konularında sıklıkla kullanılmaktadır. Doğrusal diskriminant analizi sınıf içi frekansların eşit olmaması durumunda konuyu kolayca ele alabilmektedir. Temel bileşenler analizinden farkı özelliklerin sınıflandırılmasından ziyade doğrusal diskriminant analizinde verinin sınıflandırılmasındadır. Doğrusal diskriminant analizinde iki farklı yaklaşım bulunmaktadır. Bunlardan birisi sınıf bağımlı değişim, diğeri ise sınıf bağımsız değişimdir. Sınıf bağımlı değişim yaklaşımında sınıflar arası varyans ile sınıf içi varyans oranı maksimize edilmektedir. Sınıf bağımsız yaklaşımda

ise toplam varyans ile sınıf içi varyans oranı maksimize edilmektedir. Bu yaklaşımda her sınıf diğer sınıflara göre ayrı birer sınıf olarak düşünülür. Matematiksel olarak doğrusal diskriminant analizinin işlem safhaları şu şekilde gerçekleşecektir:

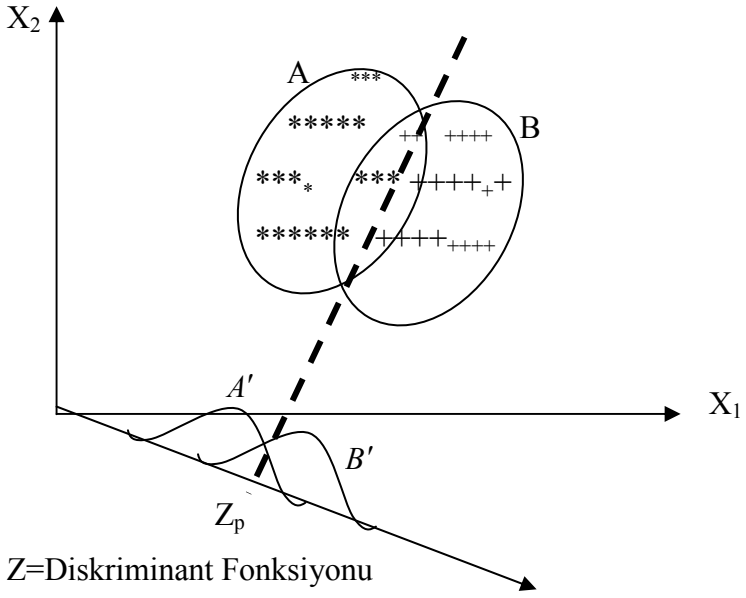
- Veri kümeleri ve test kümeleri düzenlenir. Verilen veri kümeleri ve test vektörleri formüle edilir.
- Her veri kümesinin ve bütün veri kümesinin ortalamaları hesaplanır. Örnek olarak μ_1 ve μ_2 ortalaması olan iki veri kümesinin toplam verisinin ortalaması μ_3 olsun. $\mu_3 = p_1 x \mu_1 + p_2 x \mu_2$ (1.13) olacaktır. Burada iki farklı küme olduğundan p_1 ve p_2 =.50 olacaktır.
- Sınıf ayrımının yapılabilmesi için kriterin formüle edilmesinde sınıf içi ve sınıflar arası dağılım kullanılır. Sınıf içi dağılım her sınıfın beklenen kovaryansıdır. Dağılım ölçümlerinden sınıf içi dağılım $S_w = \sum_j p_j x(\mathrm{cov}_j)$ (1.14) eşitliğine göre yapılır. Sınıflar arası dağılım ise $S_b = \sum_j (\mu_j - \mu_3) x (\mu_j - \mu_3)^T$ (1.15) olarak tanımlanır. S_b her sınıfın ortalama vektörlerinin ortalamasının üye olduğu veri kümesinin kovaryansıdır. Doğrusal diskriminant analizinin optimizasyon kriteri sınıflar arası dağılım ile sınıf içi dağılımın oranıdır. Bu kriterin maksimize edilmesi ile dönüştürülen uzayın apsisi tanımlanmaktadır.
- Özdeğer vektörlerinin dönüşümü bir boyutlu değişken olmayan değişimin uygulandığı vektör uzayında bir alt uzay olarak tanımlanır. Sıfıra eşit olmayan özdeğer vektörlerinin oluşturduğu özdeğer vektörleri kümesi doğrusal olarak bağımsız ve dönüşümde değişmeden kalmaktadır. Böylece herhangi bir vektör uzayı özdeğer vektörlerinin doğrusal birleşimi olarak temsil edilebilir. Özellikler arasındaki doğrusal bağımlılık sıfır özdeğer değerine sahiptir. Dolayısıyla sıfıra eşit olan değerler ihmal edilir.
- Herhangi bir L-sınıf probleminde L-1 adet sıfıra eşit olmayan özdeğere sahip olunur.
- Dönüşüm tamamlandığında Öklid mesafesi veri noktalarının sınıflandırılmasında kullanılır. Bu mesafenin belirlenmesinde $dist\_n = (transform\_n\_spec)^T x - \mu_{ntrans}$ (μ_{ntrans}=dönüştürülen veri kümesinin ortalaması, n sınıf endeksi, x test vektörünü temsil etmektedir) formülü kullanılır.

- *n* mesafeleri arasında en küçük Öklid mesafesi test vektörünü *n* sınıfına ait olmasına göre sınıflandırır.

1.5.1. İki gruplu doğrusal diskriminant analizi

İki gruplu doğrusal diskriminant analizi, birimlerin çok sayıdaki değişkene göre anakütleyi birbirinden ayırma problemi üzerinde durmaktadır. İki gruplu diskriminant analizinde anakütle grupları önceden belirlendikten sonra, bu iki ana kütle ile ilgili özellikler ölçülmektedir. Bu yaklaşım birimleri sınıflandıran diğer analiz yöntemlerinden diskriminant analizini ayırmaktadır.

İki gruplu diskriminant analizi öncelikle grafik üzerinde açıklanmaktadır. Şekil 1.3, iki gruplu bir diskriminant fonksiyonunu göstermektedir. Bağımlı değişkenin iki gruplu (A ve B) ve her iki grup üzerinde ölçülen özelliklerinde iki tane (X_1 ve X_2) olduğu varsayılmaktadır. Şekilde yıldız işaretleri (*) A grubunun ve artı (+) işaretleri B grubunun birimlerini göstermektedir. Elipsler ise, ilgili gruba ait birimlerin yaklaşık %95 veya daha fazlasını içermektedir. İki elipsin kesiştiği noktaları birleştiren bir doğru çizilir ve daha sonra bu doğruyu yeni bir eksen (Z) üzerine dik olarak izdüşümü alınırsa A' ve B' tek değişkenli dağılımlarının kesişim kümesi diğer çizilebilecek doğrulara kıyasla daha küçük olmaktadır. Böylece iki grubun ayrımı en iyi şekilde sağlanmış olmaktadır (Albayrak 2006).



Şekil 1.3. İki Gruplu Diskriminant Fonksiyonu

Z eksenine diskriminant fonksiyonu veya standart normal dağılım ekseni adı verilmektedir. Z ve elipslerin kesiştiği noktaya (Z_p) kritik değer (ayırıcı değer) denilmektedir. Diskriminant fonksiyonu yardımıyla X_1 ve X_2 değişkenleri ile tanımlanan birimler bu değişkenlerin doğrusal bileşimi olarak tek bir diskriminant değerine dönüştürülmektedir. Bu nedenle her birim Z ekseni üzerinde (A' ve B' dağılımlarıyla gösterildiği gibi)bir nokta olarak gösterilebilir. Birimlerin Z ekseni üzerindeki izdüşümleri kritik noktanın (Z_p) sağında veya solunda olmasına göre, noktanın temsil ettiği birim A veya B grubuna atanmaktadır. İki gruplu diskriminant analizi için bir tane Z ekseni yeterlidir. Eğer bireyleri tanımlayan değişken veya grup sayısı ikiden fazla olması halinde olayı grafikle gösterme olanağı güçleşmektedir (Albayrak 2006).

İki gruplu diskriminant analizinde iki yaklaşım vardır. Bunlardan birisi Fisher, diğeri ise Mahalanobis yaklaşımıdır. Fisher'in yaklaşımı diskriminant puanları üzerine temellendirilmiştir. Fisher, gruplar arası azami fark yaratabilecek diskriminant skorlarının üretilebileceği bağımsız X değişkenlerinin doğrusal birleşiminin bulunmasını önermiştir.

En iyi diskriminant değerleri üretecek doğrusal birleşimin bulunması için Fisher'in "azami farklı" düşüncesini hesaplayabilecek bir hedef fonksiyonu belirlenmelidir. k ile doğrusal birleşim ifade edilirse diskriminant değerleri $t=Xk$ olacaktır. Fisher, t diskriminant değerlerinin gruplar arası kareler toplamı ile grup içi kareler toplamı oranını maksimize edecek k'yi seçmeyi önermiştir. Fisher'in oranı:

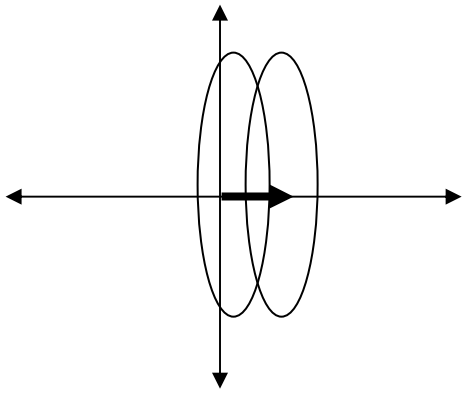
$$\frac{k'dd'k}{k'C_w k} \qquad (1.16)$$

Burada $d = (\overline{X}_{(2)} - \overline{X}_{(1)})$ vektörüdür ve iki grup ortalamaları arasındaki farkı ifade eder. C_w ise X'in grup içi kovaryans matrisidir. $\bar{t}_{(1)} = \bar{x}'_{(1)}$ ve $\bar{t}_{(2)} = \bar{x}'_{(2)}$ arasındaki farkın daha büyük olması hedef fonksiyonunun da büyümesine yol açacaktır. t diskriminant değerleri arasındaki grup içi değişim küçüldükçe hedef fonksiyonunun değeri büyüyecektir. Fisher'in belirlediği oranın azamiye çıkarılması için k şu şekilde seçilmiştir:

$$k \propto C_w^{-1} d \qquad (1.17)$$

Böylece diskriminant fonksiyonu değerleri, grup içi kovaryans matrisinin tersi ve iki grup ortalamaları arasındaki fark konularının sonuçları ile orantılıdır. k vektörünün ölçeğinin belirlenmemesi sebebi ile genellikle standartlaştırılarak seçilir. (örneğin k'nın uzunluğu 1'e eşitlenebilir.)

Bunun nasıl işlediğini görmek için bazı diskriminant fonksiyonu problemlerinden yararlanılabilir. Sadece iki bağımsız değişkeni içeren problem dikkate alındığında X_1 ve X_2 arasında grup içi kovaryans yoksa ne yapılmalıdır sorusu düşünülebilir. Bu durumda $C_w=C_w^{-1}=I$ ve k=d olacaktır. Bu durumda diskriminant fonksiyonunun apsisi iki grup ortalaması ile eşit olacaktır. Şekil 1.4 bu durumu göstermektedir.



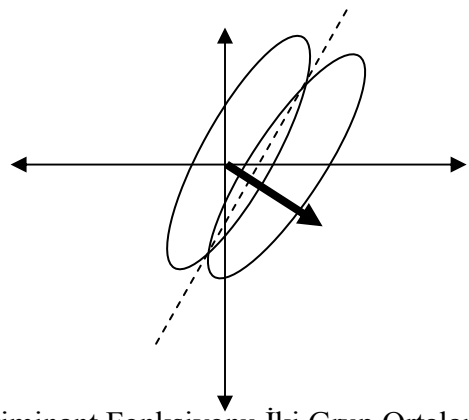
Şekil 1.4. Diskriminant Fonksiyonu İki Grup Ortalamasının Eşit Olması Durumu

Grup ortalamaları $\bar{x}'_{(1)} = (0,0)$ ve $\bar{x}'_{(2)} = (0,0)$'dır ve grup içi kovaryans matrisi ise;

$$C_w = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

olur. Diskriminant fonksiyonu değeri k, $C_w^{-1}d = (1,0)$, olarak x eksenine eşit olacaktır.

Şayet X_1 ve X_2 ilişkili ise bu durumda Şekil 1.5'e ulaşılır. Şekil 1.5'de X_1 ve X_2 arasında pozitif bir kovaryans olduğu görülebilir.



Şekil 1.5. Diskriminant Fonksiyonu İki Grup Ortalamasının Eşit Olmaması Durumu

Bu durumda grup ortalamaları aynı kalırken grup içi kovaryans matrisi şu şekilde olacaktır:

$$C_w = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Bu pozitif korelasyonun varlığı ile en iyi diskriminant fonksiyonu artık X_1 apsisi ile bulunamayacaktır. Diskriminant fonksiyon değerleri, $C_w^{-1}d$, şu şekilde oluşacaktır:

$$C_w = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Pozitif kovaryans, diskriminant fonksiyonunu aşağıya doğru çekmiştir. Apsisin değişikliği diskriminant fonksiyon değerlerinin grup içi değişimini ve grup ortalamaları için diskriminant fonksiyon değerleri arasındaki uzaklığı azaltmaktadır.

Fisher'ın Doğrusal Diskriminant Analizi ile ilgili literatürde yapılan araştırmada aşağıdaki eşitlikte belirtilen bir hedef fonksiyonu olduğu ve bu hedef fonksiyonunun maksimizasyonunun hedeflendiği belirtilmektedir (Gao ve Davis 2005; Tao vd 2005; Tang vd 2005; Tang vd 2004; Xie ve Qiu 2006; Park ve Park 2007; Zheng vd 2007; Mardia vd 1979):

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \qquad (1.18)$$

$$S_B = \sum_c N_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \qquad (1.19)$$

$$S_w = \sum_c \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T \qquad (1.20)$$

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} x_i \qquad (1.21)$$

Mahalanobis ise Fisher'den biraz daha farklı bir yaklaşım önermektedir. Diskriminant değeri hesaplamak yerine Mahalanobis uzaklıkları kullanmayı önermiştir. Birimler hesaplanan Mahalanobis uzaklığına göre yakın olduğu gruba atanmaktadır. *i* ve *k* birimleri arasındaki Mahalanobis uzaklığı aşağıdaki formülle hesaplanmaktadır:

$$MD_{ik}^2 = \frac{1}{1-r^2}\left[\frac{(x_{i1}-x_{k1})^2}{s_1^2} + \frac{(x_{i2}-x_{k2})^2}{s_2^2} - \frac{2r(x_{i1}-x_{k1})(x_{i2}-x_{k2})}{s_1 s_2}\right] \qquad (1.22)$$

Formülde s_1^2 ve s_2^2, birinci ve ikinci değişkenin varyansını; r, iki değişken arasındaki korelasyon katsayısını göstermektedir. Mahalanobis uzaklığı hesaplanırken tüm gruplara eşit ağırlıklar verilmektedir. Bu kısıtlamayı ortadan kaldırmak için Mahalanobis uzaklığı F oranına dönüştürülmektedir. F oranı hesaplanırken gruplara örnek büyüklüklerine göre ağırlık verilmektedir. Gruplar arası F oranı, grup çiftleri grup çiftleri arasındaki farklılaşmayı ölçmektedir (Sharma 1996). İlgili F oranı aşağıdaki formülle hesaplanmaktadır:

$$F = \frac{(n-1-p)(n_1 n_2)}{p(n-2)(n_1+n_2)} D_{ab}^2 \qquad (1.23)$$

1.5.2. Adımsal diskriminant analizi

İki Gruplu Diskriminant Analizinde kullanılan bir diğer teknik ise Adımsal (Stepwise) Diskriminant Analizidir. Şimdiye kadar yapılan analizlerde diskriminant fonksiyonunu ayırıcı değişkenlerin bilindiği varsayılmaktaydı. Ancak gerçekte durum böyle olmamaktadır. Potansiyel değişkenler bilinmesine rağmen bu değişkenlerden hangilerinin en iyi ayırt edici değişken olduğu bilinmemektedir. Bu amaçla adımsal diskriminant analizi kullanılmaktadır.

Diskriminant fonksiyonunu en iyi şekillendirecek değişken seti, geriye doğru (backward), ileriye doğru (forward) ve adımsal (stepwise) seçim yöntemleriyle de belirlenmektedir. Bu yöntemler değişken seçiminde kullanılan ölçütler dışında çoklu regresyon analizine benzemektedir.

Adımsal seçim yönteminde ileriye ve geriye doğru seçim beraber kullanılmaktadır. Yöntem diskriminant fonksiyonunda değişken olmadan çözümlemeye başlamakta ve her adımda fonksiyona sadece bir değişken alınmakta veya çıkartılmaktadır. İlk adımda, belirlenen giriş ölçütüne göre en iyi ayırıcı değişken modele alınmaktadır. Birinci değişken modele alındıktan sonra, modelin dışında kalan değişkenler belirlenen ölçüte göre yeniden değerlendirilmekte ve kabul edilebilirlik açısından en iyi değere sahip değişken modele ikinci değişken olarak seçilmektedir. Yine bu aşamada birinci adımda modele alınan değişken, belirlenen modelden çıkarma ölçütüne göre yeniden değerlendirilmekte ve modelden çıkarma ölçütünü karşılıyorsa bu değişken modelden çıkartılmaktadır.

Adımsal diskriminant analizinde seçim kriteri olarak Wilk's Lamda (Wilk's Lambda), F oranı (Partial F Ratio), Rao's V ve Mahalanobis kareli uzaklık ölçütü kullanılmaktadır.

Wilk's λ değeri, SS_w / SS_t oranına eşittir. Her adımda modeldeki diğer değişkenlerin etkisi yok edildikten sonra en küçük λ değeri F istatistiğine dönüştürülebildiği için en büyük kısmi F değerine sahip değişken modele alınmaktadır. Bu sayede λ değerinin minimize edilmesi ile eşzamanlı olarak SS_w değeri minimize edilmekte ve SS_b değeri maksimize edilebilmektedir. Yani λ ölçütüne göre değişken seçimi grup içi homojenliğe ve gruplar arası farklılaşmaya dayanmaktadır. F istatistiğindeki değişim ise şu formülle hesaplanmaktadır:

$$F_D = \left(\frac{n-g-p}{g-1} \right) x \left(\frac{1-\lambda_{p+1}/\lambda_p}{\lambda_{p+1}/\lambda_p} \right) \qquad (1.24)$$

F_D , F istatistiğindeki değişimi (kısmi F değerini); λ_p, modele değişken alınmadan önceki λ değerini; λ_{p+1}, modele değişken ilave edildikten sonraki λ değerini; g, örnekteki grup sayısını ve n, birim sayısını göstermektedir.

Rao'nun V istatistiği literatürde Lawley-Hotelling Trace olarak da bilinir. Rao'nun V istatistiği aşağıdaki formülle hesaplanmaktadır:

$$V = (n-g) \sum_{i=1}^{p} \sum_{j=1}^{p} w^*_{ij} \sum_{k=1}^{g} \left(\overline{X}_{ik} - \overline{X}_i \right) x \left(\overline{X}_{jk} - X_j \right) \qquad (1.25)$$

Formülde $\overline{X}_{ik}$, i'nci değişkenin k grubu ortalamasını; $\overline{X}_i$, i'nci değişkenin ortalamasını; w_{ij}^*, gruplar arası kovaryans matrisinin tersini; n_k, k grubunun birim sayısını; g, modeldeki grup sayısını ve p, modeldeki değişken sayısını göstermektedir. Grup ortalamaları arasındaki daha büyük fark daha büyük V değeri demektir. Modele değişken alınıp çıkartılırken V istatistiğindeki değişim, *p(g-1)* serbestlik derecesiyle χ^2 dağılımına uymaktadır. Her ne kadar V istatistiği, grup farklılıklığını maksimize ediyorsa da grup homojenliğini dikkate almamaktadır. Bu nedenle V istatistiğiyle türetilecek diskriminant fonksiyonu grup içi maksimum homojenliğe sahip değildir (Albayrak 2006; Klecka 1980; Sharma 1996).

p değişkenli g grup arasındaki karesel uzaklığı veren ve Mahalanobis tarafından ileri sürülen D^2 uzaklığı aşağıdaki gibi hesaplanır:

$$D_{ij}^2 = (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j) \qquad (1.26)$$

D^2 uzaklığının i ve j gruplarını birbirinden ayırmada etkin rol oynayıp oynamadığı ise Hotelling T^2 yaklaşımı ile test edilebilir.

$$T^2 = \left(\frac{n_1 x n_2}{n_1 + n_2} \right) D^2 \qquad (1.27)$$

T^2'nin önemliliğinin belirlenmesi için de F yaklaşımından yararlanılır.

$$F = \frac{(n_1 + n_2 - p - 1)}{p(n_1 + n_2 - 2)} T^2 \qquad (1.28)$$

F'nin önemliliği, *p, (n_1+n_2-p-1)* serbestlik dereceli F dağılımının kritik değerleri kullanılarak belirlenir.

Modeldeki değişkenler çıkarma ölçütüne göre değerlendirildikten sonra modelin dışındaki değişkenler modele giriş için yeniden değerlendirilmektedir. Böylece değişkenler modelden çıkarma ölçütünü sağladıkları sürece modele alınmaktadır. Modelin giriş ve çıkış ölçütlerini sağlayan değişken kalmadığı zaman değişken seçim işlemine son verilmektedir.

Bir diğer olasılık kriteri ise gruplar arasındaki artık varyansın (residual variance) azaltılmasını amaçlamaktadır. Formülü şu şekildedir:

$$R = \sum_{i=1}^{g-1} \sum_{j=i+1}^{g} \frac{4}{4 + D^2 (G_i | G_j)} \qquad (1.29)$$

Toplamdaki tüm terimler, önceden belirlenen sınıflara ait değişkenler ile ayırt edici değişkenler arasındaki çoklu kanonik korelasyon değerinin karesinin birden çıkarılmasının tahminidir. Bu artık varyanstır, çünkü tüm terimler ayırt edici değişkenler tarafından açıklanamayan kukla değişken içerisindeki değişim oranıdır. R grupları eşit olarak bölmeye çalışır.

1.5.3. İkiden çok gruplu diskriminant analizi

Diskriminant analizi ikiden fazla grup oluşturmak için de kullanılmaktadır. Çoklu diskriminant analizinin amacı, iki gruplu diskriminant analizi ile aynıdır. Ancak, iki gruplu diskriminant analizinde gruplar arasındaki farkları gösterebilmek için tek bir diskriminant fonksiyonu yeterli iken çoklu diskriminant analizinde gruplar arası farkları tanımlayabilmek için bir veya daha çok sayıda fonksiyon türetilebilmektedir. Kısaca, çoklu diskriminant analizinde gruplar arasındaki farklılaşmayı sağlayacak minimum diskriminant fonksiyonu sayısının belirlenmesi gerekmektedir (Albayrak 2006).

İki gruplu diskriminant analizinde olduğu gibi yine Fisher'in ve Mahalanobis'in yaklaşımları incelendiğinde mantıksal olarak iki gruplu analize göre bir fark bulunmadığı görülmektedir.

A (π_1) ve B (π_2), ……., G(π_g) isimli G tane yığın olsun. Bu yığınlardan n birimlik p tane birbirleri ile ilişkili gözlem yapılsın. X veri matrisinde gözlemler yığından yığına az ya da çok farklılık gösterir. X matrisi, π_1 yığınında alınan örnekler için X_1 gözlem matrisi, π_2 toplumundan alınan örnekler için X_2 gözlem matrisi ve π_G yığınından alınan örnekler için X_G gözlem matrisi elde edilir. Bu yığınlar için olasılık fonksiyonları $f_1(X)$, $f_2(X)$, ……$f_G(X)$ olacaktır. Bu veri matrisinden örnek ortalama vektörleri ve kovaryans matrisleri aşağıdaki şekilde hesaplanır:

$$\bar{x}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} x_{1j} \; ; \; s_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)' \qquad (1.30)$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j} \; ; \; s_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)(x_{2j} - \bar{x}_2)' \qquad (1.31)$$

İncelenen yığınların aynı kovaryans matrisine sahip oldukları kabul edilerek, örnek kovaryans matrisleri S_1 ve S_2' nin birleşimi S_p (pooled variance-covariance matrix) aşağıdaki şekilde hesaplanır:

$$S_{pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)} \qquad (1.32)$$

Ortak kovaryans matrisi, $\hat{\Sigma} = E(X - \bar{x}_i)(X - \bar{x}_i)'$ biçiminde de hesaplanabilir. X gözlem matrisi kullanılarak elde edilen çok değişkenli gözlem matrisi tek değişkenli *y* değerlerine dönüştürülür. Bu *y* değerleri X gözlem matrisinin doğrusal bileşenleridir.

Doğrusal diskriminant analizinde her bir grup için birer diskriminant fonksiyonu aşağıdaki şekilde hesaplanır:

$$Y_i = b_{0i} + b_{1i}X_1 + b_{2i}X_2 + ... + b_{pi}X_p \quad \text{i=1,2,..(grup sayısı)} \quad (1.33)$$

Bu fonksiyonda b_{0i} sabit değeri, b_{ij} ise doğrusal bileşenleri belirtmektedir. Doğrusal bileşenlere kanonik değişkenler adı da verilmektedir.

Her bir grup için değişkenlere ilişkin doğrusal bileşenler, değişkenlerin diskriminant fonksiyonundaki etkinliklerini, belirleyiciliklerini göstermektedir.

Doğrusal bileşenler, $b_{ij} = S^{-1}(\bar{x}_i)$ i=1,2,..g; j=1,2,…p biçiminde hesaplanır. b_i katsayılarına göre grup diskriminant fonksiyonu katsayıları ya da kanonik değişkenler adı verilir. Bazen katsayılar ölçeklendirilerek ya da standartlaştırılarak kullanılmaktadır. Standartlaştırmanın amacı katsayıları genel katsayılar içinde ağırlıklandırarak elemanların kolay yorumlanması sağlanır.

Gruplar arası farkı maksimize edecek bir diskriminant fonksiyonu aracılığı ile grupları birbirinden ayırmak mümkün olacaktır. Bu nedenle ortak bir ayırma fonksiyonu belirlenir. i ve j grupları arasındaki ayırma fonksiyonu;

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_pX_p \qquad (1.34)$$

şeklinde yazılır. Bu fonksiyondaki b_i doğrusal bileşenleri de ortalama fark vektörü aracılığı ile aşağıdaki gibi bulunur:

$$b_i = S^{-1}(\bar{x}_i - \bar{x}_j) \quad \text{i=1,2,….,g} \qquad (1.35)$$

Sabit değer olan b_0 katsayısı ise $b_0 = -(1/2)\bar{x}'S^{-1}\bar{x}$ (1.36) şeklinde hesaplanır. Her gruba ilişkin diskriminant fonksiyonları ise aşağıdaki gibi hesaplanır:

S ortak kovaryans matrisi ve $\bar{x}_i$ i 'inci grup ortalama vektörü olmak üzere her bir grubun b_i katsayılar vektörü hesaplanır.

$$b_{ij} = S^{-1}\bar{x}_i \quad \text{i=1,2,…..,g; j=1,2,……,p} \qquad (1.37)$$

Sabit değer ise $b_{i0} = -(1/2)\bar{x}_i' S^{-1}\bar{x}_i$ (1.38) şeklinde hesaplanır.

Gruplara göre belirlenen sabit ve kanonik katsayılar değişken değerleri ile çarpılarak doğrusal diskriminant fonksiyonları belirlenir.

Mahalanobis yaklaşımında ise her grubun sentroide olan uzaklığı bulunmaya çalışılmakta ve her gruba bu sentroidlere yakınlığı ölçüsünde atama yapılmaktadır. Probleme katılan G grup için mesafe hesaplaması şu eşitlik kullanılarak yapılmaktadır:

$$D_g^2 = (x - \bar{x}_{(g)})' C_w^{-1} (x - \bar{x}_{(g)}) \qquad (1.39)$$

Burada x' 'nün g grubuna ait olma ihtimali $P(g|x')$ ise Bayes teoremi ile şu şekilde formüle edilmiştir (Lattin vd 2003):

$$P(g|x') = \frac{q_g P(x'|g)}{q_1 P(x'|1) + ... + q_G P(x'|G)} \qquad (1.40)$$

1.5.4. Fisher'ın doğrusal diskriminant fonksiyonu

X; p değişkenlerinin rasgele *px1* vektörü ve varyans-kovaryans matrisi Σ, kareler toplamı matrisi ise *T* olarak kabul edilsin. γ px1 vektörünün ağırlıkları olarak alındığında diskriminant fonksiyonu (1.41)'deki eşitlikte olduğu gibi olacaktır:

$$\xi = X'\gamma \qquad (1.41)$$

Diskriminant skorlarının sonuçları için kareler toplamı ise;

$$\begin{aligned} \xi'\xi &= (X'\gamma)'(X'\gamma) \\ &= \gamma' X X' \gamma \\ &= \gamma' T \gamma \quad \text{olacaktır.} \end{aligned} \qquad (1.42)$$

T grup içi ve gruplar arası kareler toplamlarının toplamı olduğundan (T=B+W) 1.42'deki eşitlik şu şekilde yazılabilir:

$$\begin{aligned} \xi'\xi &= \gamma'(B+W)\gamma \\ &= \gamma' B \gamma + \gamma' W \gamma \end{aligned} \qquad (1.43)$$

Diskriminant analizinin amacı ağırlık vektörünün ,γ, tahmin edilebilmesidir. Dolayısıyla 1.41'deki eşitlik kullanılarak

$$\frac{2(B\gamma - \lambda W\gamma)}{\gamma' W\gamma} = 0$$

$$\begin{aligned}(B - \lambda W)\gamma &= 0 \\ (W^{-1}B - \lambda I)\gamma &= 0\end{aligned} \qquad (1.44)$$

olarak bulunur. Buradan da

$$\left|W^{-1}B - \lambda I\right| = 0 \qquad (1.45)$$

genel çözümü elde edilmiş olur. $W^{-1}B$ simetrik olmayan matrisin özdeğer vektörünün bulunmasına yardımcı olmaktadır. Özdeğer vektörleri ile de diskriminant fonksiyonu için ağırlık matrisi elde edilmiş olur. İki gruplu durum için ise 1.44'deki eşitlik daha da basitleştirilebilir. İki grup için B değeri,

$$\begin{aligned}B &= \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)(\mu_1 - \mu_2)' \\ &= C(\mu_1 - \mu_2)(\mu_1 - \mu_2)' \qquad (1.46)\end{aligned}$$

şeklinde hesaplanabilir. μ_1 ve μ_2 grup 1 ve grup 2'nin px1 vektörlerinin ortalamalarını; n_1 ve n_2 ise grup 1 ve grup 2'deki gözlem sayılarını ifade etmektedir. C ise sabit sayıdır. Buradan yola çıkılarak eşitlik 1.44 şu şekilde yazılabilir:

$$\begin{aligned}\left[W^{-1}C(\mu_1 - \mu_2)(\mu_1 - \mu_2)' - \lambda I\right]\gamma &= 0 \\ CW^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)'\gamma &= \lambda\gamma \\ \frac{C}{\lambda}\left[W^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)'\gamma\right] &= \gamma \qquad (1.47)\end{aligned}$$

$(\mu_1 - \mu_2)\gamma$ teriminin sayıl (scalar) olması sebebi ile eşitlik 1.47 şu şekilde yazılabilir:

$$\gamma = KW^{-1}(\mu_1 - \mu_2) \qquad (1.48)$$

burada $K = C(\mu_1 - \mu_2)'\gamma / \lambda$ olduğundan sayıldır ve bu sebeple de sabit olarak kabul edilir. Grup içi kovaryans matrislerinin eşitliği varsayımından yola çıkılarak eşitlik 1.48 şu şekilde de yazılabilir:

$$\gamma = K\Sigma^{-1}(\mu_1 - \mu_2) \qquad (1.49)$$

Şayet K sabitinin 1 değerine eşit olduğunun farz edersek eşitlik 1.49 şu şekle dönüşür:

$$\gamma = \Sigma^{-1}(\mu_1 - \mu_2)$$

veya

$$\gamma' = \Sigma^{-1}(\mu_1 - \mu_2)' \qquad (1.50)$$

Eşitlik 1.50 ile verilen fonksiyon Fisher'ın diskriminant fonksiyonunu temsil etmektedir. K değerinin farklı değerler almasının γ değerlerini değiştireceği açıktır.

1.5.5. Kuadratik diskriminant analizi

Doğrusal diskriminant fonksiyonunun normallikten uzaklaşmayı engellemede kuvvetli, fakat eğik dağılımlarda kullanılamayacağı bilinmektedir. Bu varsayımların bozulduğu durumlarda alternatif fonksiyonlar kullanılır. Kuadratik diskriminant fonksiyonu verilerin normal dağıldığı ancak grupların varyans-kovaryans matrislerinin farklı olmaları durumunda kullanılan fonksiyondur. Kovaryans matrislerinin eşitliği varsayımı nadiren görülebilen bir durumdur (Lachenbruch 1975: 20).

Kuadratik diskriminant analizinde katsayıların hesaplanmasında ortak kovaryans matrisi yerine (S) grupların kovaryans matrislerinin farkları alınır.

$$Q(x) = \frac{1}{2}\log\frac{|S_j|}{|S_i|} - \frac{1}{2}(x^{-(i)}S_i^{-1}x^{-(i)} - x^{-(j)}S_j^{-1}x^{-(j)} + x(S_i^{-1}x^{-(i)} - S_j^{-1}x^{-(j)})) - \frac{1}{2}x(S_i^{-1} - S_j^{-1})x \quad (1.51)$$

Başlangıçta iki grup için geliştirilen bu fonksiyon ikişerli alınarak çok grup olma durumu için de kullanılır. Fonksiyonda S_i ve S_j sırasıyla i'nci ve j'nci gruba ilişkin varyans-kovaryans matrisleridir. $S_i=S_j=S$ alınırsa; kuadratik fonksiyon doğrusal fonksiyona eşit olacaktır.

Fonksiyon değeri $Q(x) \geq 0$ ise bireyin R_i bölgesine, değilse R_j bölgesine sınıflandığı bu yöntemde, hatalı sınıflandırma olasılığı:

$$R_{Q(x)} = \left[1 + \exp(Q(x) - \log(\hat{q}_j / \hat{q}_i))\right]^{-1} \qquad (1.52)$$

eşitliği ile ifade edilir.

Kovaryans matrislerinin eşit olmaması durumunda bir önceki işlemlere ilave olarak ($\Sigma_1 \neq \Sigma_2$) ise sınıflandırma bölgeleri R_1 ve R_2 şu şekilde hesaplanmaktadır:

$k = \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\mu_1{}'\Sigma_1^{-1}\mu_1 - \mu_2{}'\Sigma_2^{-1}\mu_2)$ olmak üzere,

$$R_1 = -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu_1{}'\Sigma_1^{-1} - \mu_2{}'\Sigma_2^{-1})x - k \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (1.53)$$

$$R_2 = -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})x - k < \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (1.54)$$

'dir. Sınıflandırma bölgeleri x'in kuadratik fonksiyonu olarak tanımlanmaktadır. Kovaryans matrislerinin eşit olması durumunda $\Sigma_1 = \Sigma_2$ olacağından $-\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x$ kuadratik terimi yok olacaktır ve sınıflandırma bölgeleri kovaryans matrislerinin eşitliğinde olduğu gibi hesaplanabilecektir.

Şayet π_1 ve π_2 yığınları çok değişkenli normal yoğunluk fonksiyonuna sahiplerse ve ortalama ve kovaryans matrisleri $\mu_1;\Sigma_1$ ve $\mu_2;\Sigma_2$ olarak kabul edilirse x_0'ın π_1 yığınına tahsis edilmesi şayet,

$$-\frac{1}{2}x_0'(\Sigma_1^{-1} - \Sigma_2^{-1})x_0 + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})x_0 - k \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (1.55)$$

şartı sağlanırsa yapılabilecektir. Aksi takdirde x_0, π_2 yığınına tahsis edilecektir.

1.6. Parametrik Olmayan Diskriminant Analizi Teknikleri

Parametrik testlerin varsayımlarından kaynaklanan kısıtların dikkate alınmaması için parametrik olmayan diskriminant analizi teknikleri kullanılmaktadır. Parametrik olmayan diskriminant analizi teknikleri genel olarak Esnek Diskriminant Analizi (Flexible Discriminant Analysis) olarak adlandırılmaktadır. Nominal, Ordinal, Aralık ve Oran ölçeklerinde bile ayırma ve gruplandırma işlemleri yapılabilmektedir. Esnek Diskriminant Analizi kendi içerisinde yer alan ve sınıflama işlevini yerine getiren model ya da prosedür adı verilebilecek yöntemlerden oluşmaktadır. Bunlardan en yaygın kullanıma sahip olanları Parametrik Olmayan Regresyon, Optimal Skorlara Dayalı Regresyon, Esnek Diskriminant Analizi, Cezalandırılmış Diskriminant Analizi (Penalized Discriminant Analysis), Karma Diskriminant Analizi(Mixture Discriminant Analysis), Çok Değişkenli Uyarlanmış Regresyon Splinleri (Multivariate Adaptive Regression Splines-MARS), BRUTO Yöntemi v.b. isimlerle anılan yöntemleri içermektedir (Öztürk 2006).

Özellik ayırt etme perspektifi ile diskriminant analizi iki kare matris olan S^E ve S^I'nın temel olarak alındığı bir tekniktir. Bu matrisler genel olarak farklı sınıflar

arasında (S^E) ve sınıf içi (S^I) örneklem vektörlerinin dağılımını temsil eder. Birkaç kriter bu matrislerin sınıf dağılımını ölçmek için tek bir istatistiğe dönüştürülmesi maksadıyla teklif edilmiştir. Bu ölçümler özellik seçimi ve özellik ayırt etmede kullanılmaktadır. Fukunaga ve Mantock (1983) parametrik diskriminant analizi ile ilgili sınırlamaların üstesinden gelebilmek maksadıyla parametrik olmayan bir metot ortaya koymuşlardır.

Parametrik olmayan diskriminant analizinde sınıflar arası dağılma matrisi diğer sınıfta bölgesel olarak işaret edilen vektörlerden alınır. Bu ekstra sınıf en yakın komşu $x \in C_k$ için şu şekilde tanımlanmaktadır:

$$x^E = \left\{ x' \in \overline{C}_k \,/ \left\| x' - x \right\| \le \left\| z - x \right\|, \forall z \in \overline{C}_k \right\} \tag{1.56}$$

Aynı şekilde sınıflar arası en yakın komşu matrisi de ifade edilebilir:

$$x^I = \left\{ x' \in L_c \,/ \left\| x' - x \right\| \le \left\| z - x \right\|, \forall z \in \overline{C}_k \right\} \tag{1.57}$$

İki tanım da *k*-En Yakın Komşu durumuna genişletilirse x^E ekstra veya sınıflar arası örneklerin ortalaması olacaktır. Bu durumda parametrik olmayan sınıflar arası dağılım matrisi şu şekilde tanımlanır:

$$S^E = \frac{1}{N} \sum_{n=1}^{N} w_n (x - x^E)(x - x^I)^T \tag{1.58}$$

Yapılan çalışmalar göstermiştir ki elde edilen denklem Fisher'ın Diskriminant Analizinin genişletilmiş bir durumudur. Ayrıca düşünülen komşu sayısı toplam uygun olan örneklem sayısına eşitlendiğinde ayırt edilen özellikler parametrik olmayan diskriminant analizinde bulunan ile Fisher'ın Diskriminant Analizinde bulunana eşit olmaktadır (Bressan ve Vitria 2003).

1.6.1. Esnek diskriminant analizi

Doğrusal diskriminant analizi ile hesaplanan diskriminant fonksiyonlarının katsayıları, gerçek gözlemlerin birer doğrusal bileşeni olan doğrusal bir eşitlik ile regresyon formunda ifade olarak alınabilir. Doğrusal diskriminant analizi yaklaşımı parametrik olmayan vektör uzayına genişletilmiş gizli bir doğrusal regresyon yöntemi olarak ele alınır ve yeniden düzenlenirse bu yönteme esnek diskriminant analizi denir. Başka bir açıdan doğrusal diskriminant analizinin Kernelize edilmek suretiyle genelleştirilmiş hali olarak ta esnek diskriminant analizi ifade edilebilir.

Esnek diskriminant analizi doğrusal olmayan sınırların kullanımına izin vermektedir. Esnek diskriminant analizinde diskriminant kriteri şu denklemle gösterilmektedir:

$$ASR = \frac{1}{n}\sum_{l=1}^{L}\ \sum_{i=1}^{n}(\theta_1(g_i) - x_i'\beta_l)^2 \qquad (1.59)$$

Denklemde ASR artık kare ortalamasını (Averaged Squared Residual), g_i i'inci örneklemin sınıfını, x_i ise i'inci örneklem için tahmin edici p'lerin vektörünü temsil etmektedir (Reynes vd 2006).

Ancak esnek diskriminant analizi bu kadarla bitmemektedir. $x_i'\beta_l$ aslında denklemin doğrusal regresyon olduğuna işaret etmektedir. Bu durumda bu terim değiştirilmek suretiyle denklem parametrik olmayan regresyon haline dönüştürülebilir. Dönüşüm için $\eta_l(x_i)$ terimi parametrik olmayan regresyon terimi olarak kullanılmıştır. Bu sayede doğrusal diskriminant analizinin doğrusallık eksikliği de giderilmiş olacaktır. Başlangıçta belirtilen kriter ise artık şu şekle dönüşmüş olacaktır:

$$ASR = \frac{1}{n}\sum_{l=1}^{L}\ \sum_{i=1}^{n}(\theta_1(g_i) - \eta_l(x_i))^2 \qquad (1.60)$$

Esnek diskriminant analizinde kullanılacak bu diskriminant kriteri ile doğrusallık varsayımı taşımayan verilerin kullanılması ile sınıflandırma yapılabilecektir.

1.6.2. Cezalandırılmış diskriminant analizi

Cezalandırılmış Diskriminant Analizi, doğrusal diskriminant analizi yaklaşımının net olarak belirlenmemiş, kabaca ortaya konmuş koordinatlar içerisinde sınırlandırılmış olan ve temelde bir regresyon yöntemi olarak ele alınan bir yaklaşımdır. Bu yaklaşımda katsayıların belirli sınırlandırmalar (penalized) ile düzleştirilmesi (smoothing) yapılarak uzayda uygun formların oluşturulması sağlanır.

Cezalandırılmış diskriminant analizinde çok fazla korelasyona sahip tahmin edicilerin olması durumunda bu değişkenlerin düşük performansa yol açmaları engellenmektedir. Bunun için ise doğrusal diskriminant analizinde gruplar arası kovaryans matrisi Σ_w yerine $\Sigma_w + \lambda\Omega$ terimi kullanılmaktadır. Burada Ω ceza matrisini simgelemektedir (Hastie vd 1995).

1.6.3. Karma diskriminant analizi

Karma Diskriminant Analizi her bir sınıfın Gaussian karışım yaklaşımı içinde yer almasına izin veren çok sınıflı prototip sınıfların oluşturulmasını sağlayan bir yaklaşımdır. Karma Diskriminant Analizi yaklaşımında her bir sınıf Gaussian karışım yaklaşımı içinde gösterilerek çoklu sınıf prototipleri belirlenir. Bu üçüncü durumda ise farklı merkezler (centroids) ile iki ya da daha fazla Gaussian karışım yaklaşımı tarafından her bir sınıfın modellemesidir, fakat her parçanın (competent) aynı kovaryans matrisinde sınıflar içinde ve sınıflar arasında her bileşenin Gaussian karışım yaklaşımı olmak şartı aranır. Bu genişleme Karma diskriminant analizi olarak adlandırılır.

Karma diskriminant analizinde giriş aşamasında ağırlıkların belirlenmesi ve ağırlıkların maksimizasyonu için 4 aşamalı bir işlem yürütülmektedir. Bu aşamalar şu şekildedir:

- Başlangıç Adımı: $\hat{\mu}_{jr}$ (ortalamaların başlangıç tahmini), $\hat{\pi}_{jr}$ (karma olasılıklar) ve $\hat{\Sigma}$ (ortak kovaryans matrisi) parametrelerinin başlangıç tahminlerinin alınması (k-yönlü kümeleme analizi sonucuna göre)
- Beklenti Adımı: Ağırlıklar hesaplanır.

$$\hat{p}(c_{jr}|x,j) = \frac{\pi_{jr} e^{-D(x\mu_{jr})/2}}{\sum_{r=1}^{R_j} \pi_{jr} e^{-D(x\mu_{jr})/2}} \tag{1.61}$$

- Maksimizasyon Adımı: Ağırlıklandırılmış karma olasılıkların, ortalamanın ve kovaryans matrisinin hesaplanması

$$\hat{\pi}_{jr} \alpha \sum_{g_i=j} p(c_{jr}|x_i,j), \sum_{r=1}^{R_j} \hat{\pi}_{jr} = 1 \tag{1.62}$$

$$\hat{\mu}_{jr} = \frac{\sum_{g=j} x_i p(c_{jr}|x_i,j)}{\sum_{g=j} p(c_{jr}|x_i,j)} \tag{1.63}$$

$$\hat{\Sigma} = (1/N) \sum_{j=1}^{J} \sum_{g_i=j} \sum_{r=1}^{R_j} p(c_{jr}|x_i,j)(x_i - \mu_{jr})(x_i - \mu_{jr})^T \tag{1.64}$$

- Son Adım: Adım 2 ve 3'ün $\hat{p}(c_{jr}|x,j)$ 10^{-10}'dan daha fazla değişmeyinceye kadar tekrar edilmesi

Bu adımlar neticesinde elde edilen değerler kullanılarak sınıflandırmanın yapılabilmesi için sonsal(aposteriori) olasılıklar hesaplanmaktadır. Sonsal sınıf olasılıkları Bayes teoremi kullanılarak hesaplanmaktadır:

$$\Pr(G=j|X=x)=\frac{\Pi_j \Pr(x|j)}{\Pr(x)}=\frac{\Pi_j \sum_{r=1}^{R_j}\pi_{jr}e^{-D(x\mu_{jr})/2}}{\sum_{j=1}^{J}\Pi_j \sum_{r=1}^{R_j}\pi_{jr}e^{-D(x\mu_{jr})/2}} \qquad (1.65)$$

Bayes teoremi ile belirtilen eşitlikte $\Pr(x|j)$ sınıf şartlı yoğunluğu ve Pr(x) ise şartsız yoğunluğu temsil etmektedir. Yeni x nesnesinin en büyük sonsal olasılıkla sınıfa ataması yapılmaktadır. Karma diskriminant analizinin diğer olasılık yaklaşımlarına göre en önemli faydası homojen olmayan ve kümelenmiş verilere uygulanabilmesidir (Schmid vd 2009).

İKİNCİ BÖLÜM
2. LOJİSTİK REGRESYON ANALİZİ

Lojistik Regresyon Analizi kategorik verileri analiz etmeye yarayan ve araştırmalarda sıklıkla kullanılan bir yöntemdir. Sosyal Bilimlerde yapılan araştırmalardan sağlık bilimlerinde yapılan araştırmalara, ekonomiden pazarlama ve bankacılık alanına kadar çok geniş bir alanda ilişkisel analiz yapılmasına olanak tanır.

Lojistik Regresyon Analizinin yaygın bir şekilde kullanılmaya başlanması ile katsayı tahmin yöntemleri daha fazla geliştirilmiş ve lojistik regresyon modelleri daha detaylı bir şekilde incelenmeye başlanmıştır. Cornfield (1962) lojistik regresyondaki katsayı tahmin işlemlerinde diskriminant fonksiyonu yaklaşımını ilk kez kullanarak popüler hale getirmiştir. Lee (1984) basit dönüşümlü (cross-over) deneme planları için lineer lojistik modeller üzerinde durmuştur. Bonney (1987) lojistik regresyon modelinin kullanımı ve geliştirilmesi üzerinde çalışmış olup Robert ve diğerleri (1987) ise lojistik regresyonda standart ki-kare, olabilirlik oran (G^2), en çok olabilirlik tahminleri, uyum iyiliği ve hipotez testleri üzerinde çalışma yapmışlardır. Duffy (1990) lojistik regresyonda hata terimlerinin dağılışı ve parametre değerlerinin gerçek değerlere yaklaşımını incelemiştir.

Çok değişkenli istatistiksel verilerin sınıflandırılmasında kullanılan çok değişkenli istatistiksel yöntemlerden biri olan lojistik regresyon analizinde verilerin yapısındaki grup sayısı bilinmekte ve bu verilerden hareketle bir ayrımsama modeli oluşturulmaktadır (Ulupınar 2007:39).

Lojistik regresyon analizinde diskriminant analizinde belirtilen varsayımların olmaması ve bağımsız değişkenlerin kategorik olabilmesi bu tekniğin kullanımını kolaylaştırmaktadır.

Lojistik regresyon analizinin temel amacı diğer regresyon yöntemlerinde olduğu gibi bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi incelemektir. Başka bir deyişle amaç minimum uygun sayıda değişken ile sonuç değişkeni ve açıklayıcı değişkenler arasındaki ilişkiyi tanımlayan kabul edilebilir modeli kurmaktır. Lojistik regresyon yönteminde bağımlı değişkenin sürekli olması gibi bir varsayım yoktur, özellikle bağımlı değişkenin iki veya daha çok değer aldığı durumlarda kullanılır (Ulupınar 2007:39).

Lojistik regresyon analizi, bağımlı değişkenin türüne göre 3 farklı şekilde kullanılabilir:

- İkili (Binary) Lojistik Regresyon,
- Sıralı (Ordinal) Lojistik Regresyon,
- İsimsel (Nomial ve Multinomial) Lojistik Regresyon.

İkili lojistik regresyon yönteminde sınıflayıcı değişken iki sonuçludur. Bu değişken sayısal veya kısa alfanümerik bir değişken olabilir. Analizde sınıflayıcı değişken bağımlı değişken olarak referans kabul edilir ve bağımsız değişkenlerle olan ilişkisi incelenerek sınıflandırmada kullanılacak tahmini regresyon denklemi kurulur. Kurulan denklem yardımıyla sınıfların tahminine çalışılır.

Sıralı lojistik regresyon bağımlı değişkenin üç veya daha fazla cevaplı olması durumunda uygulanan bir yöntemdir. Ayrıca cevaplar arasında sıralı (ordinal) bir ilişki de olması gerekir.

İsimsel lojistik regresyon yöntemi ise Sıralı lojistik regresyona benzer ancak burada bağımlı değişkenin aldığı cevapların sıralı olması şartı aranmamaktadır.

Elde edilen gözlem değerlerine lojistik regresyon analizi uygulanacağına karar verildikten sonra katsayıların tahmini, yorumlanması, katsayılara ilişkin hipotez testlerinin yapılması ve modelin başarısının değerlendirilmesi gerekmektedir.

Lojistik regresyon analizinde bağımlı değişkene bağlı olarak bağımsız değişkenlerin dağılımının doğrusal olmadığı durumlarda kullanılabilir olduğu görülmektedir (Tabachnick ve Fidell 2001:517). Bağımsız değişkenlerin doğrusal olmaması sebebi ile eşitliğin ifadesi çoklu regresyon modellerinden biraz daha karmaşık olabilmektedir. Logit modeller;

$$\ln(\frac{P}{1-P}) = A + \sum B_j X_{ij} \qquad (2.1)$$

eşitliği ile ifade edilmektedir.

2.1. Lojistik Regresyon Analizinde Değişken Seçimi

Lojistik Regresyon Analizi; sürekli, kesikli, ikili ya da bunların bir karışımı olan veri setlerinden kategorik bir sonucu tahmin etmeye olanak sağlar. Lojistik regresyon modellerinde kategorik bağımsız değişken/değişkenler, sadece sürekli bağımsız

değişken/değişkenler veya hem kategorik hem de sürekli bağımsız değişkenler kullanılabilir (Kaşko 2007:19).

Bağımlı değişkendeki değişimi açıklayabilmek için kurulan bir regresyon eşitliğine girecek değişken sayısı ne kadar çok olursa, eşitlik o kadar küçük hata taşımaktadır. Ancak, gerek bağımsız değişkenlerin birisiyle gözlem elde etmenin getireceği yük, gerekse bu gözlemleri belirli bir zaman aralığında yapma mecburiyetinin getireceği zorluklar ve olası hatalar bağımsız değişken sayısını azaltmayı zorunlu kılabilir. Bu nedenle, sınıflandırma tahmininin doğruluğu mümkün olduğunca yüksek tutulmalı; ayrıca ekonomik yük ve zorlukların yanı sıra, fazla değişkenle ilgili veri elde etmenin getirebileceği sistematik hataları mümkün olduğunca azaltabilecek sayıda bağımsız değişkenle çalışılması araştırıcılar açısından önemli bulunmaktadır (Düzgüneş ve diğerleri 1987).

Lojistik regresyon analizinde değişken seçimi analize bağımsız değişkenin nasıl dâhil edileceği ile ilgilidir. Farklı yöntemler kullanılarak değişkenlerin seçimi yapılabilmektedir. Diğer çok değişkenli yöntemlerde olduğu gibi adımsal seçim modellerinde bir sonraki aşamada hangi değişkenin modele dâhil edilebileceğine karar verilmektedir. İstatistiksel olarak algoritmalardan hiçbirisi en iyi modeli sağlamayı garanti edememektedir. Bu aşamada farklı modellerin denenip bu modeller arasından sınıflandırma başarısına göre seçim yapmak en iyi yaklaşım olmaktadır (Albayrak 2006).

Kullanılan yaklaşımlar ise şunlardır:

- Tüm Değişkenlerin Modele Dâhil Edildiği Yaklaşım: Bütün değişkenler bir blok olarak tek aşamada modele dâhil edilir.
- İleri Seçim (Koşullu): İleriye doğru adımsal bir yöntemdir. Değişkenler modele teker teker alınarak kriterleri sağlamayanlar modelde tutulmaz. Değişkenler modele alınırken skor istatistiğinin önemine, çıkarılırken de koşullu parametre tahminlerine dayanan olabilirlik oranına göre karar verilir.
- İleri Seçim (Olabilirlik Oranı): İleriye doğru adımsal bir yöntemdir. Değişkenler modele alınırken skor istatistiğinin önemine, çıkarılırken de maksimum kısmi olabilirlik tahminlerine dayanan olabilirlik oranına göre karar verilir.
- İleri Seçim (Wald): İleriye doğru adımsal bir yöntemdir. Değişkenler modele alınırken skor istatistiğinin önemine, çıkarılırken de Wald istatistiğine göre karar verilir.

- Geriye Eleme (Şartlı): Geriye doğru adımsal bir seçim yöntemidir. Önce tüm değişkenler modele alınır daha sonra birer birer kriterleri sağlamayan değişkenler modelden çıkartılır. Tüm geriye doğru yöntemlerde önce tüm değişkenler alınıp sonra teker teker çıkarma yaklaşımı geçerlidir. Değişkenler modelden çıkarılırken koşullu parametre tahminlerine dayanan olabilirlik oranına göre karar verilir.
- Geriye Eleme (Olabilirlik Oranı): Geriye doğru adımsal seçim yöntemidir. Değişkenler modelden çıkarılırken kısmi olabilirlik tahminlerine dayanan olabilirlik oranına göre karar verilir.
- Geriye Eleme (Wald): Geriye doğru adımsal seçim yöntemidir. Değişkenler modelden çıkarılırken Wald istatistiğine göre karar verilir (SPSS Regression Models 16.0).

2.2. İkili (Binary)Lojistik Regresyon Modeli

Çeşitli gösterim biçimleri olan genel doğrusal regresyon modeli,

$$E(y_{\cdot} / x_{i1},...,x_{ip}) = \sum_{k=0}^{p} \beta_k x_{ik}; i = 1,...,n \text{ için} \qquad (2.2)$$

biçiminde koşullu beklenen değer olarak da yazılması mümkündür. Bu modelde açıklayıcı değişkenler üzerinde kısıt yok iken, y bağımlı değişkeninin sürekli olması koşulu vardır. Herhangi bir i'inci gözlem için,

$$y_i = \sum_{k=0}^{p} \beta_k x_{ik} + u_i \qquad (2.3)$$

biçiminde ifade edilen modelde açıklayıcı değişkenler üzerinde bir kısıt olmadığından y_i sonuç değeri $-\infty$ ile $+\infty$ arasında tüm değerleri alabilmektedir. Bağımlı değişkenin 0,1 gibi değerler aldığı durumda bu kural bozulmakta ve $P(y_i=1)$, i 'inci gözlemin 1 değerini alma olasılığı olmak üzere, beklenen değer,

$$E(y_i) = P(y_i = 1) = \sum_{k=0}^{p} \beta_k x_{ik} \qquad (2.4)$$

olarak bulunur. Sol tarafı 0-1 arasında değerleri alan bu denkleme doğrusal olasılık modeli adı verilmektedir (Tatlıdil 1996:290).

Açıklayıcı değişkenlerin sınırsız değerler alması nedeniyle söz konusu eşitlik her zaman sağlanamamaktadır. Bu sebeple çeşitli dönüşümler yapılmaktadır. Bu dönüşümlerden en yaygın olarak kullanılan iki tanesi logit ve probit dönüşümlerdir.

Logit dönüşümde doğrusal olasılık modelinde olasılık değerleri üzerinde P/(1-P) dönüşümü yapılarak sonuç değişkeninin sınırları 0, $+\infty$ yapılmakta, daha sonra ise bu oran değerinin doğal logaritması alınarak sonuç değişkenin sınırları $-\infty$, $+\infty$ yapılmaktadır. Bu dönüşümlerden sonra elde edilen yeni fonksiyon,

$$E(y_i) = L_i = \log(P_i/(1-P_i)) = \sum_{k=0}^{p} \beta_k x_{ik} \tag{2.5}$$

olarak yazılmaktadır. Lojistik model ya da kısaca logit olarak bilinen bu modelde P_i olasılık değeri,

$$P_i = \exp(\sum_{k=0}^{p} \beta_k x_{ik})/(1+\exp(\sum_{k=0}^{p} \beta_k x_{ik})) \tag{2.6}$$

biçiminde tanımlanmakta ve lojistik fonksiyon adını almaktadır. Bu modelde sonuç değişkeninin iki değer alması nedeni ile hata terimi sıfır ortalama ve P(1-P) varyanslıdır. Hata terimi bu parametrelerle binom dağılımlı olup, analiz bu teorik temele dayanmaktadır.

Logit fonksiyonu aynı zamanda şu şekilde de gösterilmektedir:

$$P_i = \frac{1}{1+e^{-(\alpha+\beta x)}} \tag{2.7}$$

Bu eşitliğe lojistik dağılım fonksiyonu adı verilir. $\alpha+\beta x=Z$ olarak kabul edilirse bu durumda,

$$\frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i} \tag{2.8}$$

eşitliğine ulaşılır. Bu eşitlik odds (bahis) oranı olarak adlandırılır. Odds oranı daha özet bir ifadeyle olayın gerçekleşme olasılığının olayın gerçekleşmeme olasılığına olan oranını ifade etmektedir. Odds oranından genellikle ikili değişken arasındaki ilişkinin ölçümlenmesinde yararlanılır. Etki katsayısı veya etki büyüklüğü olarak tanımlanan Exp(β), aynı zamanda Odds oranını vermektedir ve bu değer açıklayıcı değişkenlerin etkisinin kolayca yorumlanabilmesi açısından önemlidir.

Odds oranının doğal logaritması alınırsa Logit'e ulaşılır. Yani odds oranının logaritması katsayı tahminleri bakımından yalnız X'e göre değil ana kütle katsayılarına

göre de doğrusaldır (Gujarati 2001:555). Ayrıca odds oranları, x'in arttığı her birim için e^{β}'nın katları kadar artar.

Böylece odds oranının logaritması alınmak suretiyle doğrusal olmayan ilişki logit fonksiyonu yardımıyla doğrusal hale getirilmiştir.

2.3.Logit Modelin Özellikleri

Logit modeller normal dağılım, kovaryans matrislerinin eşitliği gibi kısıtlayıcı varsayımlara sahip olmadığından diğer yöntemlere göre avantaja sahiptir. Ayrıca bağımlı değişkenin kesikli olması yöntemin uygulanabilirliği üzerinde bir etki yaratmamaktadır. Son olarak model parametreleri logaritmik odds oranları kullanılarak kolayca izah edilebilir ve yorumlanabilir (Hosmer ve diğerleri 1991:1630).

2.4. Modelin Parametre Tahmini

Modelin katsayılarının tahmininde En Çok Olabilirlik Yöntemi, Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi, Minimum Logit Ki-Kare Yöntemi kullanılmaktadır. Açıklayıcı değişkenlerin hepsi sürekli ise minimum logit ki-kare yöntemi, değişkenlerin hepsi kesikli ise en çok olabilirlik yöntemi, hem sürekli hem de kesikli ise ağırlıklandırılmış iteratif en küçük kareler yöntemi kullanılmaktadır (Başarır 1990:12-13).

2.4.1. En çok olabilirlik yöntemi

Lojistik regresyon çözümlemesinde bağımlı değişken ile bağımsız değişkenler arasındaki ilişki doğrusal olmadığı için model parametreleri en küçük kareler yöntemi ile tahmin edilemez. Başarı olasılığı $P_i=P(y_i=1/x)$, başarısızlık olasılığı $1-P_i$ olduğunda i'inci gözlem için olasılık,

$P(y_i / x_i) = P_i^{y_i}(1-P_i)^{1-y_i}; i = 1,.....,n$ (2.9) için

biçiminde yazılacak olursa, bu olasılık n gözlem için olabilirlik fonksiyonu olarak,

$$L(y/x) = P(y/x) = \sum_{i=1}^{n} P_i^{y_i}(1-P_i)^{1-y_i} \qquad (2.10)$$

biçiminde ifade edilebilir. Bilindiği gibi en çok olabilirlik yöntemi p açıklayıcı değişkene ilişkin β'ların kestirimini, sonuç değişkeni y'nin gözlenme olabilirliğini maksimum kılacak biçimde bulmayı amaçlamaktadır. Yani L(y/x,β) olabilirlik fonksiyonunu maksimum yapacak $\hat{\beta}$ katsayılar vektörünü belirlemek ana hedeftir. Bu durumda yukarıdaki eşitliklerden yararlanılarak lojistik modelin olabilirlik fonksiyonunun logaritması,

$$\log L(y/x,\beta) = \sum_{i=1}^{n} (y_i \log P_i + (1-y_i)\log(1-P_i)) \qquad (2.11)$$

biçiminde olup, bunun β'ya göre birinci türevi,

$$\sum_{i=1}^{n} (y_i - P_i)x_{ij} = 0; j = 1,...., p \qquad (2.12)$$

için olabilirlik denklemini vermektedir. Bu denklemin çözümünde ise $\hat{\beta}$ kestirim değerleri bulunmaktadır. Logit modelde gösterilen P_i'nin β'larda doğrusal olmaması nedeniyle en çok olabilirlik yönteminden iteratif yolla çözüme gidilir. İteratif çözümlemede β'lara herhangi bir başlangıç değerleri verilerek elde edilen kestirimlerden, her adımda δ kadar eksiltme ya da artırma yapılıp türevler alınarak sonuca ulaşılır. Sonuca ulaşmanın göstergesi yakınsamanın sağlanmasıdır. Yakınsama ise iterasyonlar arasında fark olmaması durumunda sağlanmaktadır.

2.4.2. Yeniden ağırlıklandırılmış iteratif en küçük kareler yöntemi

Gruplandırılmış verilerde *J* grubunun her birinde n_j denemeden r_j başarı elde edildiğinde başarı oranı $P_j=r_j/n_j$ olarak tanımlanabilir. $Var(r_j/n_j)=P_j\ (1-P_j)/n_j$ olduğundan, her binom dağılımlı gözlem için varyans değişmektedir. Bu durumda logit (r_j/n_j)'nin açıklayıcı değişkenler üzerinde $w_j=n_j/P_j(1-P_j)$ ağırlığı ile ağırlıklandırılmış regresyon uygulanmalıdır. Ancak w_j ağırlık değerleri de P_j'nin bir fonksiyonu olduğu için en küçük kareler yöntemi iteratif olarak uygulanacak ve ağırlık değerleri her adımda (kestirim değerlerine bağlı olarak) yeniden elde edilecektir (Tatlıdil 1996:296).

2.4.3. Minimum logit ki-kare yöntemi

Ağırlıklı en küçük kareler kestirim yönteminin özel bir biçimi olan ve Berkson tarafından geliştirilen bu yöntemde *2xJ* çapraz tablolarındaki beklenen ve gözlenen logit değerleri arasındaki farktan yararlanılmaktadır. Yöntem tekrarlı veriler olması durumunda kullanılmaktadır. Bir önceki yöntemde verilen P_j olasılığı üzerinden yapılan logit dönüşümü, bu yöntemde sonuç değişkenini oluşturmaktadır. Kestirimde kullanılan ağırlık değerleri $n_jP_j(1-P_j)$ olarak elde edilmektedir. Bu bilgiler ışığında yöntem logit değeri olarak tanımlanan sonuç değişkeninin, açıklayıcı değişkenler ile (tanımlanan ağırlık değerleri ile ağırlıklandırılmış) regresyonundan en küçük kareler kestirimlerini elde etmeye dayanmaktadır. Buradan tek adımda bulunan ağırlıklı en küçük kareler kestirimleri minimum logit ki-kare kestirimleri adını almaktadır.

2.5. Modelin Katsayılarının Testi ve Yorumlanması

Modelin verilere uyumunun belirlenmesindeki önemli adımlardan biri, uyumun iyiliği diğer bir deyişle, modelin gözlenen verileri ne kadar iyi tanımlanabildiğinin incelenmesidir (Hosmer ve diğerleri 1991: 1610). Bağımsız değişkenlerin modele eklenmesi veya çıkarılması ile ilgili olarak yapılan analitik çalışma burada ele alınacaktır. Bu analiz ile modelde kullanılacak katsayıların önem kontrolü yapılmış olacaktır.

2.5.1. Olabilirlik oran testi

Doğrusal regresyonda regresyon kareler toplamı ne kadar büyük olursa bağımsız olmakla birlikte bu yöntemde gözlemlenen değerlerin tahmin edilen değerlerle karşılaştırılması log olabilirlik ile yapılır.

Burada $H_0 : \beta_1 = 0$ hipotezi test edilmektedir. Geçerli model sadece önemli olan değişkenleri içeren model, doymuş model ise değişken sayısı kadar parametre içeren model olmak üzere;

$$D=-2\ln(\text{Geçerli Modelin Benzerliği/Doymuş Modelin Benzerliği}) \qquad (2.13)$$

olarak hesaplanır. Parantezin içerisindeki ifade benzerlik ya da olabilirlik oranı (likelihood ratio) olarak ifade edilir ve aşağıdaki test istatistiği elde edilir.

$$D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \qquad (2.14)$$

Bir değişkenin modeldeki etkisini ölçmek için değişken modelde yer alırken ve modelden çıkartıldığında elde edilen D değerleri arasındaki farka bakılır.

G=D(Değişken İçermeyen Model)-D(Değişken İçeren Model)=-2ln(Değişken İçermeyen Modelin Olabilirliği/Değişken İçeren Modelin Olabilirliği) (2.15)

şeklinde bulunur. Burada bulunan G değeri Ki-kare dağılımına uymaktadır (Tatlıdil 1996:297).

2.5.2. Wald testi

Wald istatistiği β parametresi ile standart hatasının oranıdır ve Z dağılımı göstermektedir. Doğrusal regresyondaki t testinin alternatifidir. Wald χ^2 istatistiği t değerlerinin karesine eşittir.

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \qquad (2.16)$$

Wald istatistiği standart normal bir değişkendir. Karesi 1 serbestlik derecesi ile χ^2 dağılır. Bu durumda;

$$W = \left[\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}\right]^2 \qquad (2.17)$$

istatistiği tanımlanabilir. Büyük β değerleri için tahmin edilen standart hatalar da büyük çıkmaktadır. Bu durum H_0 hipotezi yanlış iken kabul edilmesi olasılığını artırmaktadır.

Modelin katsayılarının yorumlanması için lojistik regresyon modelinde aşağıdaki eşitlikten yararlanılır.

$$\ln\left(\frac{P_i}{1-P_i}\right) = d = X'\beta \qquad (2.18)$$

$\left(\frac{P_i}{1-P_i}\right)$'nin logaritması modelin katsayılarının doğrusal bir şekilde yorumlanabilmesini sağlamaktadır. Bu şekilde açıklayıcı değişkendeki 1 birim değişmenin olasılık üzerindeki etkisi görülmektedir (Ulupınar 2007:49).

2.5.3. Pearson ki-kare testi

Karl Pearson tarafından 1900 yılında bulunan ve değişik kullanım amaçları olmasına karşılık, var olan veya olması gereken frekanslar arasındaki farklılıkların anlamlılığının test edilmesi temeline dayanan bir başka testte Pearson ki-kare testidir (Canyüker ve Aşan 2005: 41).

$$r_i = \frac{y_i - \pi(x_i)}{\sqrt{\pi(x_i)(1-\pi(x_i)}} \qquad (2.19)$$

formülü ile hesaplanan bu istatistiğinin değerinin büyük olması yani anlamlı çıkmaması modelin verilere uyumunun başarısız olduğunu göstermektedir. Ki-kare dağılımına uyduğu ifade edilse de bazı şartlar sağlanmadıkça tam bir uyum ölçütü olarak bu istatistiğin kullanılamayacağı düşünülmektedir. Conover (1999: 241) istatistiğin ki-kare dağılımına uyması için Koehler ve Larntz'ın

- toplam gözlem sayısının $n \geq 10$
- sınıf sayısının $c \geq 3$
- beklenen değerlerin hepsinin $E \geq 0{,}25$

şeklinde önerdiği koşulların sağlanması veya bir başka yol olarak çok sayıda küçük beklenen değerlerin bir araya getirilmesi gerektiğini belirtmiştir.

2.6. Modelin Uyum İyiliğinin Ölçülmesi

Katsayılarının bulunması ve önem kontrolünden sonra modelin aşağıda verilen durumlara karşı uyum iyiliğinin test edilmesi gerekmektedir.

- Logaritmik dönüşüm yerine başka bir dönüşüm daha iyi olabilir,
- Logaritmik dönüşüm uygun olsa bile modeldeki açıklayıcı değişkenlerin bir kısmı uygun olmayabilir ya da bazı etkileşim terimlerinin de modele katılması gerekebilir,
- Değişkenlerin modelde bulunması uygundur, ancak ölçek yanlış olabilir,

- Veriler arasında aykırı değer olabilir

Bu gibi durumlara karşı lojistik modelin uyum iyiliğini araştırmada kullanılan ölçütlerden önemli olanları şunlardır:

- Tüm değişkenleri içeren model ile kestirilen modele ilişkin olabilirlik oran değerlerinin farkına dayanan (hata kareler toplamına benzer) ölçütlerin ki-kare dağılacağı düşüncesinden hareketle, kurulan modelin geçerliliği sınanmaktadır. Bu yolla modele girecek açıklayıcı değişkenlere ve eklenecek karesel terimlere karar verilmektedir.
- Hata terimlerinin, x değerlerine ya da olasılık değerlerine karşı çizimi ile aykırı değer araştırması yapılmaktadır.
- Hata kareler toplamı ve olabilirlik oranına dayalı R^2 türü ölçütler de modelin uyumunu test etmede kullanılmaktadır.
- Lojistik model ayrımsama amacıyla kullanıldığında modelin doğru sınıflandırma oranı da bir uyum iyiliği ölçütüdür.

Lojistik modellerin uyum iyiliğinin belirlenmesinde aşağıdaki hipotez testi kullanılmaktadır.

H_0 :Model uygundur

H_a : Model uygun değildir.

Bu hipotezde H_0 hipotezinin kabul edilmesi modelin anlamlı olacağını göstermektedir. Lojistik modellerde normallik varsayımının bulunmaması sebebi ile parametrik testler kullanılmamakta Ki-kare ve G^2 gibi parametrik olmayan ölçütlerden yararlanılmaktadır.

Ki-kare ve G^2 bilinen en basit parametrik olmayan ölçütlerdir. Çünkü O-gözlenen, E-beklenen değerleri, OlogO ve OlogE sırasıyla gözlenen ve beklenen olabilirlikleri göstermek üzere bu ölçütler,

$$\chi^2 = \sum \frac{(O-E)^2}{E} \qquad (2.20)$$

$$G^2 = -2\sum O\log(E/O) = 2\sum O\log(O/E) \qquad (2.21)$$

$$= -2\left(\sum O\log(E) - \sum O\log(O)\right) = -2(\log(E) - \log(O)) \qquad (2.22)$$

biçiminde tanımlanmaktadır. Tekrarlı veriler için ki-kare ölçütü; $\hat{P}_i$ olasılık kestirimi değeri olmak üzere benzer biçimde tanımlanmaktadır.

$$\chi^2 = \sum_{i=1}^{J} \frac{(y_i - n_i \hat{P}_i)^2}{n_i P_i (1 - \hat{P}_i)} \qquad (2.23)$$

Lojistik regresyon analizinde, kurulan modelin önemliliğini test etmede ve bir anlamda modele girmesi gereken açıklayıcı değişkenleri belirlemede yine G^2 yaklaşımı kullanılmaktadır. Bu amaçla önerilen sapma ölçütü, doymuş model; değişken sayısı kadar parametre içeren model, kestirilmiş model; sadece önemli olduğu düşünülen değişkenleri içeren model olmak üzere,

$$D = -2 \log \frac{(GeçerliModelinOlabilirligi)}{(DoymusModelinOlabilirligi)} \qquad (2.24)$$

biçiminde tanımlanmaktadır. Olabilirlik fonksiyonunun yazılması ile,

$$D = \sum_{i=1}^{n} d_i^{\,2} = -2 \sum_{i=1}^{n} (y_i \log(\hat{P}_i / y_i) + (1 - y_i) \log((1 - \hat{P}_i)/(1 - y_i)) \qquad (2.25)$$

biçimine dönüşen sapma ölçütü, *p* modeldeki parametre sayısını göstermek üzere *(n-p)* serbestlik dereceli Ki-kare tablo değeri ile karşılaştırılmaktadır. Ayrıca sapma değeri minimum olan model en iyi model olarak bilinmektedir.

Tekrarlı veriler için sapma ölçütü ise,

$$d_j = \pm\sqrt{2}(y_j \log((y_i / n_j \hat{P}_j) + (n_j - y_j) \log((n_j - y_j) / n_j (1 - \hat{P}_j)))^{1/2} \qquad (2.26)$$

iken

$$D = \sum_{j=1}^{J} d_j^2 \qquad (2.27)$$

eşitliğinden bulunarak *(J-p-1)* serbestlik dereceli Ki-kare tablo değeri ile karşılaştırılmaktadır.

Çoklu doğrusal regresyonda katsayıların anlamlılığına ilişkin tümel F testine karşılık gelebilecek benzer bir test, lojistik regresyon analizi için de geliştirilmiştir. L_0 sadece sabit terimden oluşan modelin olabilirlik değeri, L_1 elde edilen modelin olabilirlik değeri olmak üzere,

$$C = -2 \log(L_0 / L_1) = -2(\log L_0 - \log L_1) \qquad (2.28)$$

olarak tanımlanan ölçüt *(p-1)* serbestlik dereceli Ki-kare dağılımı göstermektedir (Tatlıdil 1996:299)

Uyum iyiliği testi yaparken ayrıca Hosmer Lemeshow testi de kullanılmaktadır. Hosmer ve Lemeshow gözlenen verilerin beklenen verilere uyum iyiliğini test etmek maksadıyla 7 adet istatistik geliştirmiştir. Ancak bunlardan sadece ikisi kullanıldığından

burada C ve H ile ifade edilen bu iki istatistiğe yer verilecektir. C istatistiği verileri tahmin edilen olasılıklara göre ayırırken, H istatistiği verileri belirlenen bir kesim noktasına (cut-off point) göre ayırır. Gruplara ayrılan verilere, daha sonra ki-kare uyum iyiliği testi uygulanmaktadır. Dolayısıyla her grup için beklenen ile gözlenen değerlerin bir karşılaştırması yapılmaktadır. Hesaplanan p değeri 0,05'den büyük ise gözlenen ve beklenen değerler arasında fark olmadığını gösteren sıfır hipotezi reddedilir. Bu durum modelin mevcut veriye olan uyumunun iyi olduğunu gösterir. Bu testte aşağıdaki istatistik kullanılmaktadır.

$$G_{HL}^2 = \sum_{j=1}^{10} \frac{(O_j - E_J)^2}{E_j(1 - E_j / n_j)} \tag{2.29}$$

Belirtilen istatistikle eşit sayıda tahmin edilen olasılıklara göre 10 grup oluşturulmakta ve bu dağılımın Ki-kare dağılımına uyduğu düşünülmektedir (Başarır 1990:28).

Modelin uyum iyiliğinin belirlenmesinde R^2 istatistiklerinden de yararlanılmaktadır. Doğrusal regresyon analizindeki R^2'ye benzer şekilde yorumlanan ve kullanılan çeşitli R^2 istatistikleri mevcuttur. Modelin uyum iyiliğini belirlemede kullanılan istatistiklerden biri McFadden'in R^2 uyum iyiliği kriteridir. $-2\ln L_0$ logit modelde yer alan sabiti ve $-2\ln L_1$ modelin tüm açıklayıcı değişkenlerini içermek kaydıyla McFadden R^2 formülü şu şekilde hesaplanır:

$$McFaddenR^2 = 1 - \frac{-2\ln L_1}{-2\ln L_0} = 1 - \frac{\ln L_1}{\ln L_0} \tag{2.30}$$

R^2 istatistiği doğrusal regresyondaki belirtme katsayısına benzer, ancak onun gibi değerlendirilmez. Değişkenler arasındaki ilişkinin gücü açısından belirtme katsayısının mümkün olduğunca yüksek çıkması istenir. Aksi takdirde modelin bağımlı değişkeni açıklamada yetersiz kaldığı düşünülebilir. Oysa McFadden R^2'nin değeri genellikle düşük çıkar ve bu modelin iyi olmadığı anlamına gelmez. Yapılan uygulamalarda R^2'nin 0,2 ile 0,4 arasında değerler aldığı gözlemlenmiştir.

Bir başka R^2 istatistiği ise Cox ve Snell R^2 istatistiğidir. Cox ve Snell R^2;

$$Cox - SnellR^2 = 1 - \left[\frac{L_0}{L}\right]^{\frac{2}{I}} \tag{2.31}$$

formülü ile gösterilir ve I gözlem sayısını göstermektedir. Formülden de anlaşılacağı üzere R^2'nin en yüksek değerinin bile birden küçük çıkması diğer bir deyişle maksimum

değerinin hiçbir zaman 1'e ulaşamaması sonucun yorumlanmasında sorun yaratmaktadır.

Diğer bir test istatistiği ise Nagelkerke R^2 istatistiğidir. Nagelkerke R^2;

$$Nagel\ker keR^2 = \frac{Cox - SnellR^2}{1 - \left[\frac{L_0}{L}\right]^{\frac{2}{I}}} \tag{2.32}$$

Bu ölçümün maksimum değeri 1 olabilir ve doğrusal regresyondaki belirtme katsayısı gibi yorumlanabilir. Nagelkerke R^2, Cox-Snell R^2'den her zaman büyük ancak doğrusal regresyondaki R^2 değerinden genellikle küçük olmaktadır.

ÜÇÜNCÜ BÖLÜM
3. YAPAY SİNİR AĞLARI

Günümüzde bilgi işleme büyük ölçüde bilgisayarlar kullanılarak yapılmaktadır. Bilgisayarlar kullanılarak çok yüksek yoğunluktaki ve boyuttaki bilgi hızlı bir şekilde işlenebilmekte ve sayısal işlemler insan beyninden daha hızlı bir şekilde yürütülebilmektedir. Ancak yapısı gereği insan beyni işlem hızı olarak yavaş olmasına rağmen bilgisayarlara göre özellikle öğrenme kabiliyetinden ötürü sinir hücrelerinin de yardımıyla yüksek karmaşıklıktaki sorunları çözmede bilgisayarlardan daha süratli çalışabilmektedir.

Sinir hücrelerinin beyinde sistemli bir şekilde çalışması bilgisayarların çalışma prensibine göre daha üstün halde bulunmasını sağlamaktadır. Ancak bilgisayar sistemlerinin de çalışma prensibinin beynin çalışma sistemine uyumlu hale getirilmesi için yapılan faaliyetler sonucunda insan beyninin temel işlemci elemanları olan sinir hücrelerinin oluşturduğu ağ yapısının matematiksel olarak modellenmesine çalışılmıştır.

Beynin bütün davranışlarını modelleyebilmek için fiziksel bileşenlerin doğru olarak modellenmesi gerektiği düşüncesi ile çeşitli yapay sinir hücreleri ve ağ modelleri geliştirilmiştir. Böylece "Yapay Sinir Ağları" denen günümüz bilgisayarlarının algoritmik hesaplama yöntemlerinden farklı bir çalışma disiplini ortaya çıkmıştır (Baş 2006:6).

Yapay Sinir Ağları, beynin fizyolojisinden yararlanılarak oluşturulan bilgi işleme modelleridir. Literatürde 100'den fazla yapay sinir ağı modeli vardır. Bazı bilim adamları, beynimizin güçlü düşünme, hatırlama ve problem çözme yeteneklerini bilgisayara aktarmaya çalışmışlardır. Bazı araştırmacılar ise, beynin fonksiyonlarını kısmen yerine getiren bir çok modelleri oluşturmaya çalışmışlardır (Öztemel 2003:29).

Yapay sinir ağlarının öğrenme özelliği, araştırmacıların dikkatini çeken en önemli özelliklerden birisidir. Çünkü herhangi bir olay hakkında girdi ve çıktılar arasındaki ilişkiyi, doğrusal olsun veya olmasın, elde bulunan mevcut örneklerden öğrenerek daha önce hiç görülmemiş olayları, önceki örneklerden çağrışım yaparak ilgili olaya çözümler üretebilme özelliği yapay sinir ağlarındaki zeki davranışın da temelini teşkil eder.

3.1.Yapay Sinir Ağlarının Kullanım Alanları

Günümüzde yapay sinir ağları eksik bilgilerle çalışabilme ve normal olmayan verilere çözüm üretebilme yeteneklerinden dolayı pek çok alanda kullanılabilmektedir. Doğrusal olmayan, çok boyutlu, gürültülü, karmaşık, kesin olmayan, eksik, kusurlu, hata olasılığı yüksek veriler ve problemlerin çözümü için bir matematiksel model bulunmaması durumlarında yaygın halde yapay sinir ağları uygulamaları yapılabilmekte ve başarılı sonuçlar elde edilmektedir. Bu amaçla geliştirilen ağlar genel olarak şu fonksiyonları yerine getirmektedir:

- Probabilistik Fonksiyon Kestirimleri,
- Sınıflandırma,
- İlişkilendirme ve Örüntü Tanımlama,
- Zaman Serileri Analizleri,
- Sinyal Filtreleme,
- Veri Sıkıştırma,
- Örüntü Tanıma,
- Doğrusal Olmayan Sinyal İşleme,
- Optimizasyon,
- Zeki ve Doğrusal Olmayan Kontrol (Öztemel 2003).

3.2. Yapay Sinir Ağlarının Özellikleri

Yapay sinir ağları ağ yapısına göre farklılık taşıyabilecek özelliklere sahiptir. Ancak literatürde yapay sinir ağları şu şekilde bir genelleme yapılarak özellikleri açısından değerlendirilmektedir:

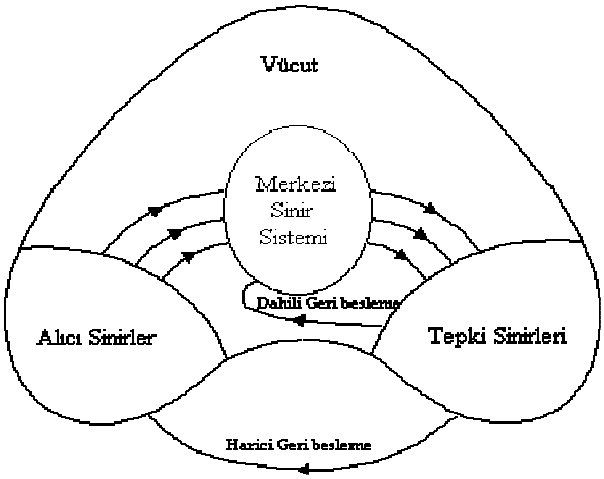
- ÖĞRENEBİLME: Klasik öğrenme algoritmalarının tersine yapay sinir ağları bir eğitim örneklem grubunu kullanarak kendi kurallarını oluşturur (Öztemel 2003:33). Öğrenmede ağ, kullanılan öğrenme tekniğine göre danışmanlı öğrenme, danışmansız öğrenme gibi tekniklerle kendi kurallarını belirlemeye çalışır. Her denek değeri ile sinaptik ağırlıkları güncelleyerek hata payını daha aza indirmeyi bu sayede hedeflemektedir. Bu ağırlıkların belirlenmesinde bazen çıktı değerlerinin istenen sonucu

vermemesi sebebi ile tekrar ayarlanması gerekebilir. Buradaki sorun hangi ağırlıkların hataya neden olduğunu bulmaktır.

- KENDİ KENDİNİ ORGANİZE EDEBİLME: Yapay sinir ağları kendi koyduğu kurallar ile çalıştığından yeni durumlara kısa sürede adapte olabilirler.
- GENELLENEBİLİRLİK: Yapay sinir ağlarının paralel dağılımlı yapısı ona genelleme yapabilme imkânı kazandırmaktadır. Ağ, bu yolla hiç karşılaşmadığı durumlar için de çözüm üretebilecek hale gelmektedir.
- UYARLANABİLİRLİK: Sinir ağlarının sinaptik ağırlıkları çevredeki değişikliklere adapte edecek şekilde değiştirme yetenekleri vardır. Özellikle de belli bir ortamda çalışmak üzere eğitilmiş bir sinir ağı, küçük değişikliklerle yeni çalışma çevresinin koşulları ile baş edecek şekilde yeniden eğitilebilir.
- ÇIKARSAMA YAPABİLME: Kısıtlı bilgilerden yola çıkılarak tam sonuca ulaşmayı sağlayabilecek yeteneğe sahiptir.
- HATA TOLERE EDEBİLME: Ağlar birbirine bağlı çok sayıda nörondan içerdiğinden, az sayıda nöron veya bağlantının bozulması sisteme bir zarar vermemektedir. Bilgi ağ içinde dağıtıldığından eksik bilgi ile de ağ çalıştırılabilir. Burada önemli olan ağın performansının değerlendirilmesinde eksik bilginin etkisinin incelenmesidir.

3.3.Biyolojik Nöron Yapısı

Biyolojik sinir sistemi, merkezinde sürekli olarak bilgiyi alan, yorumlayan ve uygun bir karar üreten beynin (merkezi sinir ağı) bulunduğu 3 katmanlı bir sistem olarak açıklanmaktadır. Alıcı sinirler (receptor), organizma içerisinden ya da dış ortamlardan algıladıkları uyarıları, elektriksel sinyallere dönüştürerek beyne iletirler. Tepki sinirleri (effector) ise, beyinin ürettiği elektriksel sinyalleri organizma çıktısı olarak uygun tepkilere dönüştürür. Şekil 3.1'de biyolojik sinir sisteminin blok şeması görülmektedir (Elmas 2003:30).



Şekil 3.1. Biyolojik Sinir Sisteminin Blok Gösterimi

Sinir hücreleri nöron olarak bilinir. Nöron, özellikle beyin olmak üzere sinir sisteminin temel birimidir ve başlıca üç kısımdan oluşur. Bunlar:

Gövde (cell body)

Gövdeye giren sinyal alıcı lifler (dendrit),

Gövdeden çıkan sinyal iletici lifler (axon).

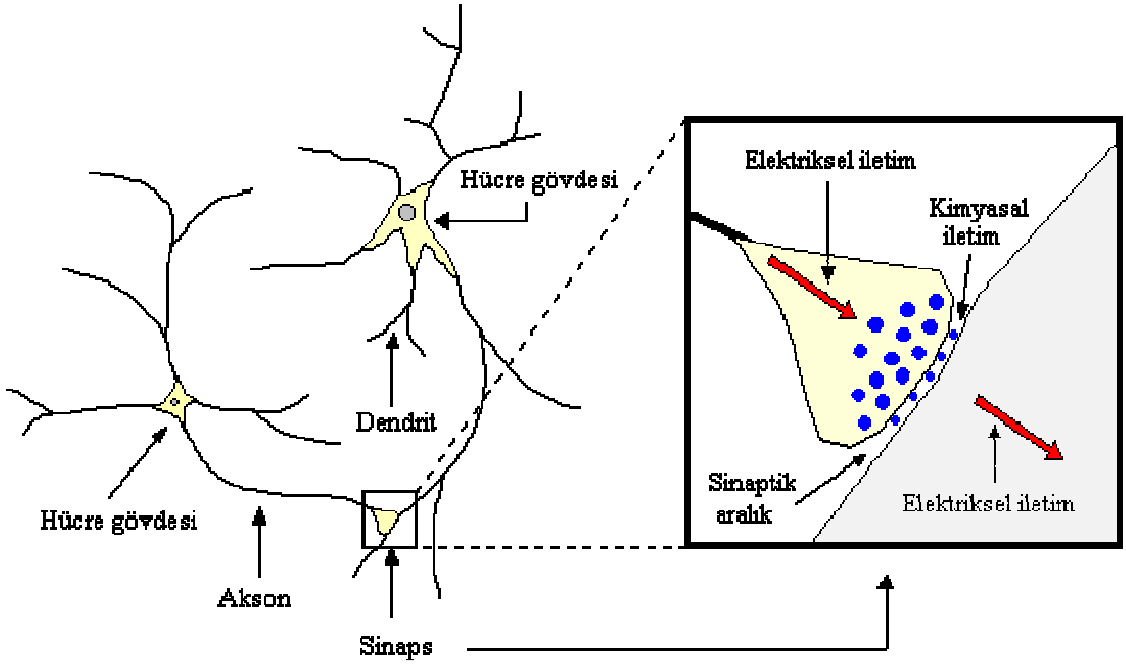
Yapay sinir ağları, insan beyninin çalışma prensibi örnek alınarak geliştirilmeye çalışılmıştır ve aralarında yapısal olarak bazı benzerlikler vardır. Bu benzerlikler Çizelge 3.1'de verilmiştir.

Çizelge 3.1. Sinir Sistemi ile Yapay Sinir Ağlarının Benzerlikleri

SİNİR SİSTEMİ	**YAPAY SİNİR AĞLARI SİSTEMİ**
Neuron	İşlem elemanı
Dendrit	Toplama/Kombinasyon fonksiyonu
Hücre gövdesi	Transfer fonksiyonu
Aksonlar	Sinir/Eleman çıkışı
Sinapslar	Ağırlıklar
Uyarıcı Giriş	Pozitif ara bağlantı ağırlığı
Engelleyici Giriş	Negatif ara bağlantı ağırlığı

Şekil 3.2'de bir nöron hücresinin yapısı verilmiştir. Bir nörondan yüzlerce, bazen de binlerce dendrit çıkabilir. Bunların uzunluğu genellikle bir milimetreden daha kısadır. Bazıları ise birkaç milimetre uzunluğa ulaşabilir. Dendritler çevre hücrelerden gelen sinyalleri (impulse) alıp, gövdeye ulaştırırlar. Her nöron, dendritler vasıtasıyla diğer birçok nöronlardan gelen işaretleri alan ve birleştiren basit bir mikro işleme birimidir.

Beyin, sıkışık olarak ara bağlaşımlı milyarlarca (yaklaşık olarak 1011) nörondan oluşmaktadır. Her eleman kendi aralarında oldukça çok sayıda nörona (eleman başına yaklaşık olarak 104 bağlantı) bağlanmıştır. Bir nöronun aksonu (çıkış yolu) ayrıştırılmıştır ve bir sinaps olarak adlandırılan bir fonksiyon vasıtasıyla diğer nöronların dendritlerine bağlanmıştır. Bu fonksiyon uçlarındaki iletim doğal olarak kimyasaldır ve işaretin miktarı, akson tarafından serbest bırakılan kimyasalların büyüklüğüne bağlı olarak transfer edilir ve dendritler vasıtasıyla alınır. Bu sinaptik büyüklük, beyin öğrenirken neyin modifiye edildiğini belirtir. Bu sinaps, beynin temel hafıza mekanizmasına dayanarak nöron içerisindeki bilginin işlenmesi ile birleştirilir (Elmas 2003:30).



Şekil 3.2. Nöron Hücresinin Yapısı

3.4.Nöron Yapısı ve Aktivasyon Fonksiyonları

Bir nöron; girdiler, ağırlıklar, toplama fonksiyonu ve aktivasyon fonksiyonundan oluşmaktadır. Girdiler elde edilen ham verilerdir ve düzenlenerek veya ham veri olarak sisteme aktarılabilmektedirler.

Ağırlıklar iki nöron arasındaki bağlantının gücünü ölçer ve bir nöronun çıktısını belirleyebilir ya da engelleyebilir. Pozitif ağırlık, bir nöronun sinyal çıkarmasını

tetiklerken, negatif ağırlık bir nöronun çıktısını engeller. Birçok sistemde ağırlıklar bilgisayar programları ile yapılır. Ağırlıkların değiştirilmesi ise sinir ağlarının öğrenmesi olarak tanımlanabilir.

Toplama fonksiyonunda genelde ağırlıklı toplam kullanılır. Ağırlıklı toplam, nöron girdilerinin sinaptik bağlantılar üzerindeki ağırlıkları ile çarpılarak bulunur. Bu fonksiyona doğrusal bağlayıcı veya toplayıcı da denilmektedir ve genellikle deneme yanılma yoluyla belirlenir (Efe ve Kaynak 2004: 6).

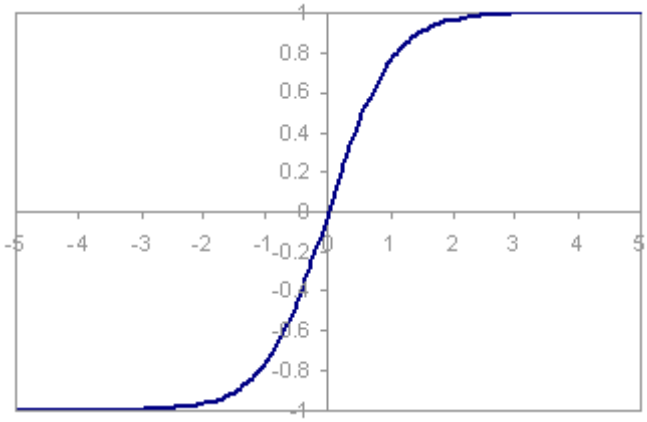
Aktivasyon fonksiyonları girdi verileri ve ağırlıklara karşılık nöronun çıktısını belirleyen matematiksel bir denklemdir. Yapay sinir ağlarında en çok tercih edilen aktivasyon fonksiyonları

- Sigmoid Fonksiyonu,
- Softmax Fonksiyonu,
- Hiperbolik Tanjant Fonksiyonu,
- Identity Fonksiyonu.

Hiperbolik Tanjant fonksiyonu şu şekildedir:

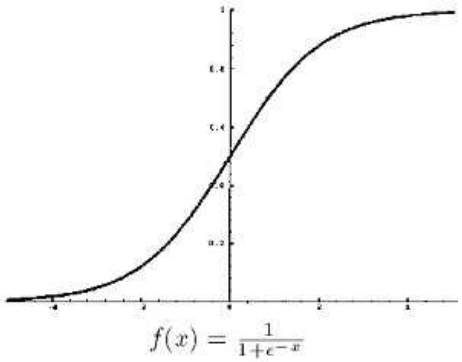
$$\gamma(c) = \tanh(c) = (e^{c} - e^{-c})/(e^{c} + e^{-c}) \qquad (3.1)$$

Hiperbolik tanjant aktivasyon fonksiyonu gerçek değerleri alarak onları, (-1,+1) aralığında olacak şekilde dönüştürür. Fonksiyon eğrisi Şekil 3.3'te gösterilmiştir.



Şekil 3.3. Hiperbolik Tanjant Aktivasyon Fonksiyonu

Sigmoid fonksiyonu ise $\gamma(c) = 1/(1+e^{-c})$ (3.2) formülünü kullanır. Sigmoid fonksiyonu da gerçek değerleri alarak (0,1) aralığındaki değerlere dönüştürür. Sigmoid fonksiyon grafiği Şekil 3.4.'te gösterilmiştir.



Şekil 3.4. Sigmoid Aktivasyon Fonksiyonu

Softmax fonksiyonu $\gamma(c_k) = \exp(c_k) / \sum_j \exp(c_j)$ (3.3) formülünü kullanır. Softmax fonksiyonu gerçek değerlerden oluşan vektörü alarak onu (0,1) aralığındaki değerlerden oluşan yeni bir vektöre dönüştürür.

Identity fonksiyonu ise $\gamma(c) = c$ (3.4) fonksiyonunu kullanır. Bu fonksiyon gerçek değerleri alır ve dönüştürmeden tekrar bilgiyi geri iade eder. Özellikle Çıktı (Output) tabakasında bu aktivasyon kullanılmaktadır (SPSS Neural Networks Tutorial 2007) .

3.5.İşlemci Eleman (Yapay Nöron)

Bir yapay sinir ağları modelinin temel birimi, Şekil 3.5'de gösterilen işlem elemanıdır. Burada girişler dış kaynaklardan veya diğer işlem elemanlarından gelen işaretlerdir. Bu işaretler, kaynağına göre kuvvetli veya zayıf olabileceğinden ağırlıkları da farklıdır.

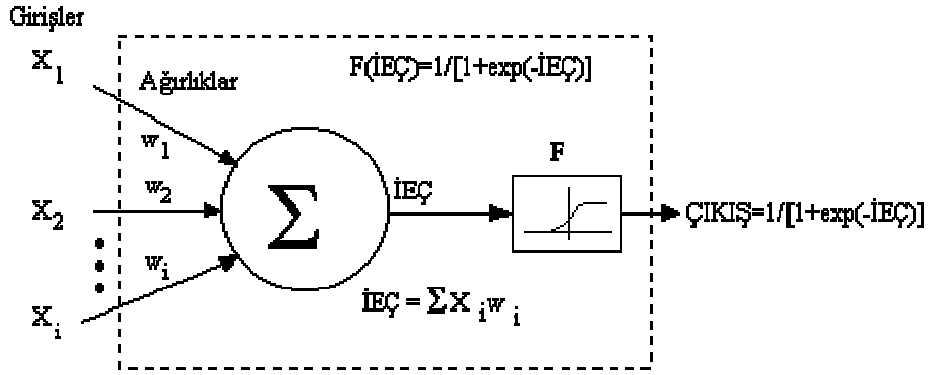
Yapay sinir ağlarında girilen giriş değerlerine önce toplama fonksiyonları uygulanır ve her bir işlem elemanının çıkış (İEÇ) değeri

$$İEÇ = \sum_{i=1}^{N} X_i W_{ij} - \theta_i \qquad (3.5)$$

olarak bulunur. Burada X_i i'inci girişi, W_{ij} j'inci elemandan i'inci elemana bağlantı ağırlığını ve θ_i eşik (threshold) değerini göstermektedir. Daha sonra bu çıkış değerleri sigmoidal aktivasyon fonksiyonuna yani öğrenme eğrisine uygulanır. Sonuçta çıkış değeri aşağıdaki şekilde bulunur.

$$ÇIKIŞ = \frac{1}{1 + e^{-İEÇ}} \qquad (3.6)$$

Transfer fonksiyonları olarak çoğunlukla, *hiperbolik tanjant* veya *sigmoid* fonksiyonu kullanılmaktadır. Şekil 3.5'de işlemci eleman çıkışında kullanılan sigmoid fonksiyona göre çıkış değerinin hesaplanması gösterilmiştir. Bu işlemci elemanın çıkış değeri diğer işlemci elemanlarına giriş veya ağın çıkış değeri olabilir (Elmas 2003:32).



Şekil 3.5. Bir İşlemci Elemanı (Yapay Nöron)

3.6. Yapay Sinir Ağlarının Sınıflandırılması

Yapay sinir ağları, genel olarak birbirleri ile bağlantılı işlemci birimlerden veya diğer bir ifade ile işlemci elemanlardan (neurons) oluşurlar. Her bir sinir hücresi arasındaki bağlantıların yapısı ağın yapısını belirler. İstenilen hedefe ulaşmak için bağlantıların nasıl değiştirileceği öğrenme algoritması tarafından belirlenir. Kullanılan bir öğrenme kuralına göre, hatayı sıfıra indirecek şekilde, ağın ağırlıkları değiştirilir. Yapay sinir ağları yapılarına ve öğrenme algoritmalarına göre sınıflandırılırlar.

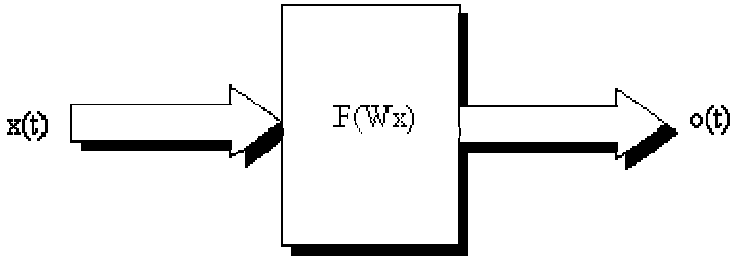
3.6.1.Yapay sinir ağlarının yapılarına göre sınıflandırılması

Yapay sinir ağları, yapılarına göre, ileri beslemeli (feedforward) ve geri beslemeli (feedback) ağlar olmak üzere iki şekilde sınıflandırılırlar.

3.6.1.1. İleri beslemeli ağlar

İleri beslemeli bir ağda işlemci elemanlar (İE) genellikle katmanlara ayrılmışlardır. İşaretler, giriş katmanından çıkış katmanına doğru tek yönlü bağlantılarla

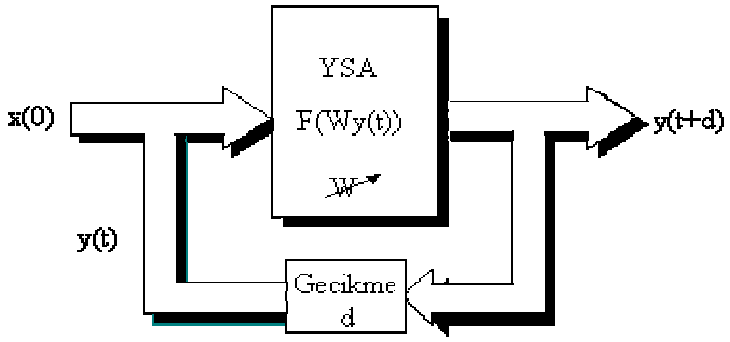
iletilir. işlemci elemanlar bir katmandan diğer bir katmana bağlantı kurarlarken, aynı katman içerisinde bağlantıları bulunmaz. Şekil 3.6'da ileri beslemeli ağ için blok diyagram gösterilmiştir. İleri beslemeli ağlara örnek olarak çok katmanlı perseptron (Multi Layer Perseptron-MLP) ve LVQ (Learning Vector Quantization) ağları verilebilir. Bu çalışmada çok katmanlı perseptron modeli kullanılacağından bu konu üzerinde durulacaktır.



Şekil 3.6. İleri Beslemeli Ağ İçin Blok Diyagram

3.6.1.2. Geri beslemeli ağlar

Bir geri beslemeli sinir ağı, çıkış ve ara katlardaki çıkışların, giriş birimlerine veya önceki ara katmanlara geri beslendiği bir ağ yapısıdır. Böylece, girişler hem ileri yönde hem de geri yönde aktarılmış olur. Şekil 3.7'de bir geri beslemeli ağ görülmektedir. Bu çeşit sinir ağlarının dinamik hafızaları vardır ve bir andaki çıkış hem o andaki hem de önceki girişleri yansıtır. Bundan dolayı, özellikle önceden tahmin uygulamaları için uygundurlar. Bu ağlar çeşitli tipteki zaman-serilerinin tahmininde oldukça başarı sağlamışlardır. Bu ağlara örnek olarak Hopfield, SOM (Self Organizing Map), Elman ve Jordan ağları verilebilir.



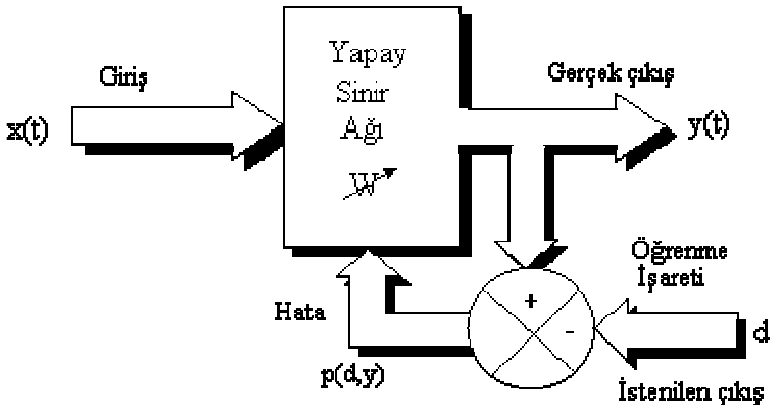
Şekil 3.7. Geri Beslemeli Ağ İçin Blok Diyagram

3.6.2.Yapay sinir ağlarının öğrenme algoritmalarına göre sınıflandırılması

Öğrenme; gözlem, eğitim ve hareketin doğal yapıda meydana getirdiği davranış değişikliği olarak tanımlanmaktadır. O halde, birtakım metot ve kurallar, gözlem ve eğitime göre ağdaki ağırlıkların değiştirilmesi sağlanmalıdır. Bunun için genel olarak üç öğrenme metodundan ve bunların uygulandığı değişik öğrenme kurallarından söz edilebilir. Bu öğrenme kuralları aşağıda açıklanmaktadır.

3.6.2.1. Danışmanlı öğrenme (Supervised Learning)

Bu tip öğrenmede, yapay sinir ağlarına örnek olarak bir doğru çıkış verilir. İstenilen ve gerçek çıktı arasındaki farka (hataya) göre işlemci elemanlar arası bağlantıların ağırlığını en uygun çıkışı elde etmek için sonradan düzenlenebilir. Bu sebeple danışmanlı öğrenme algoritmasının bir "öğretmene" veya "danışmana" ihtiyacı vardır. Şekil 3.8'de danışmanlı öğrenme yapısı gösterilmiştir. Widrow-Hoff tarafından geliştirilen delta kuralı ve Rumelhart ve McClelland tarafından geliştirilen genelleştirilmiş delta kuralı veya geri besleme (back propagation) algoritması danışmanlı öğrenme algoritmalarına örnek olarak verilebilir.

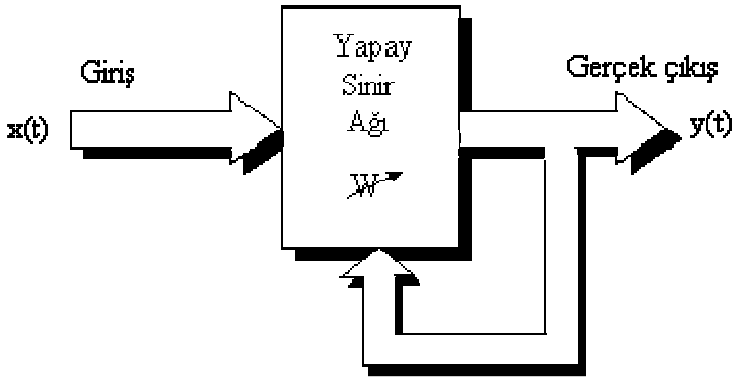


Şekil 3.8. Danışmanlı Öğrenme Yapısı

3.6.2.2. Danışmansız Öğrenme (Unsupervised Learning)

Girişe verilen örnekten elde edilen çıkış bilgisine göre ağ sınıflandırma kurallarını kendi kendine geliştirmektedir. Bu öğrenme algoritmalarında, istenilen çıkış

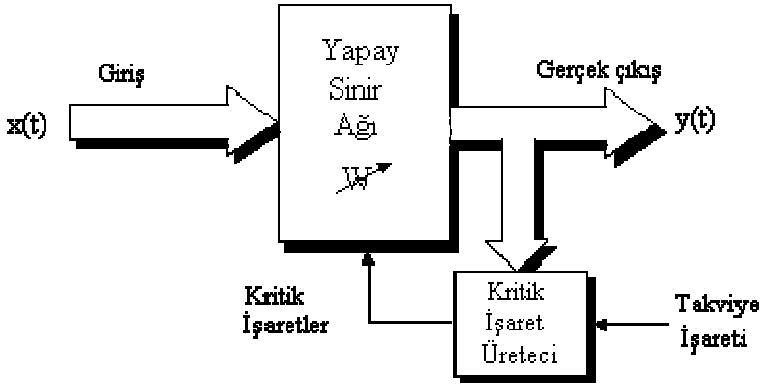
değerinin bilinmesine gerek yoktur. Öğrenme süresince sadece giriş bilgileri verilir. Ağ daha sonra bağlantı ağırlıklarını aynı özellikleri gösteren desenler (patterns) oluşturmak üzere ayarlar. Şekil 3.9'da danışmansız öğrenme yapısı gösterilmiştir. Grossberg tarafından geliştirilen ART (Adaptive Resonance Theory) veya Kohonen tarafından geliştirilen SOM (Self Organizing Map) öğrenme kuralı danışmansız öğrenmeye örnek olarak verilebilir.



Şekil 3.9. Danışmansız Öğrenme Yapısı

3.6.2.3. Takviyeli öğrenme (Reinforcement learning)

Bu öğrenme kuralı danışmanlı öğrenmeye yakın bir metottur. Denetimsiz öğrenme algoritması, istenilen çıkışın bilinmesine gerek duymaz. Hedef çıktıyı vermek için bir "öğretmen" yerine, burada yapay sinir ağlarına bir çıkış verilmemekte fakat elde edilen çıkışın verilen girişe karşılık iyiliğini değerlendiren bir kriter kullanılmaktadır. Şekil 3.10'da takviyeli öğrenme yapısı gösterilmiştir. Optimizasyon problemlerini çözmek için Hinton ve Sejnowski'nin geliştirdiği Boltzmann kuralı veya Genetik Algoritmalar takviyeli öğrenmeye örnek olarak verilebilirler.



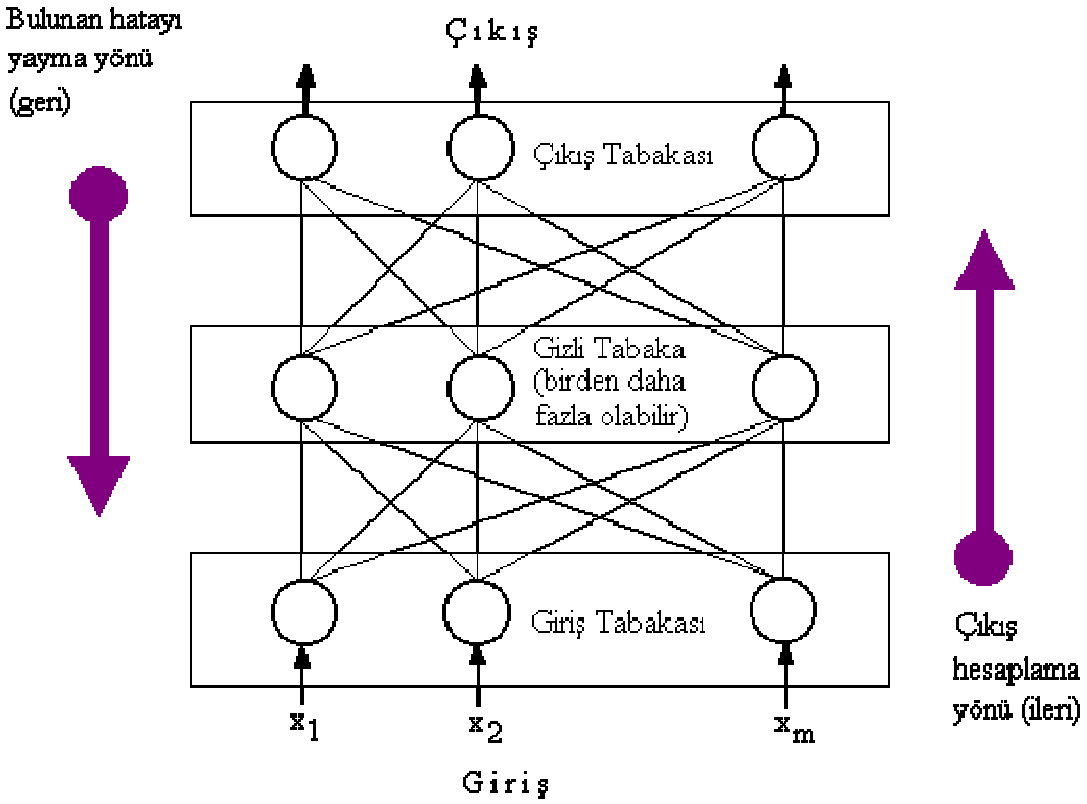
Şekil 3.10. Takviyeli öğrenme yapısı

3.7. Çok Katmanlı Perseptronlar ve Öğrenme Algoritmaları

Çok katmanlı perseptronlar (Multi Layer Perceptron-MLP)'ler birçok öğretme algoritması kullanılarak eğitilebilirler. Bu çalışmada çok gelişmiş ülkeler ile orta düzeyde gelişmiş ülkelerin SPSS 16.0 yazılımı ile çok katmanlı perseptronlar kullanılarak sınıflandırılmasına çalışılmıştır. Bu sebeple tek katlı ve çok katlı yapay sinir ağı modellerinin tamamı hakkında bilgi verilmeden doğrudan uygulamada kullanılan model izah edilmeye çalışılmıştır.

3.7.1. Çok katmanlı perseptronlar

Çok katmanlı perseptron sinir ağı modeli, Şekil 3.11'de gösterilmiştir. Bu ağ modeli özellikle mühendislik uygulamalarında en çok kullanılan sinir ağı modeli olmuştur. Birçok öğretme algoritmasının bu ağı eğitmede kullanılabilir olması, bu modelin yaygın kullanılmasının sebebidir. Bir çok katmanlı perseptron modeli, bir giriş, bir veya daha fazla ara ve bir de çıkış katmanından oluşur. Bir katmandaki bütün işlem elemanları bir üst katmandaki bütün işlem elemanlarına bağlıdır. Bilgi akışı ileri doğru olup geri besleme yoktur. Bunun için ileri beslemeli sinir ağı modeli olarak adlandırılır. Giriş katmanında herhangi bir bilgi işleme yapılmaz. Buradaki işlem elemanı sayısı tamamen uygulanan problemler giriş sayısına bağlıdır. Ara katman sayısı ve ara katmanlardaki işlem elemanı sayısı ise, deneme-yanılma yolu ile bulunur. Çıkış katmanındaki eleman sayısı ise yine uygulanan probleme dayanılarak belirlenir.



Şekil 3.11. Geri Yayılım Çok Katmanlı Perseptron Yapısı

Çok katmanlı perseptron ağlarında, ağa bir örnek gösterilir ve örnek neticesinde nasıl bir sonuç üreteceği de bildirilir (danışmanlı öğrenme). Örnekler giriş katmanına uygulanır, ara katmanlarda işlenir ve çıkış katmanından da çıkışlar elde edilir. Kullanılan eğitme algoritmasına göre, ağın çıkışı ile arzu edilen çıkış arasındaki hata tekrar geriye doğru yayılarak hata minimuma düşünceye kadar ağın ağırlıkları değiştirilir.

Çok katmanlı perseptron genelleştirilmiş delta kuralı adı verilen öğrenme kuralını kullanır. Genelleştirilmiş delta kuralı, en küçük kareler yöntemine dayalı bir öğrenme kuralı olan ve ADALINE ve tek katmanlı perseptron modellerinin öğrenme kurallarının daha gelişmiş hali olan Delta Kuralının genelleştirilmiş halidir. Delta kuralı ve genelleştirilmiş delta kuralı, danışmanlı bir öğrenme kuralıdır. Burada da temel amaç ağın beklenen çıktısı ile ürettiği çıktı arasındaki hatayı en aza indirmektir. Çok katmanlı perseptrona eğitim sırasında hem girdiler hem de o girdilere karşılık gelen çıktılar gösterilir.

SPSS 16.0 ile yapılan analizlerde ileri beslemeli MLP ağlar kullanılmıştır. Üç katmanlı yapıda giriş katmanı, gizli katman ve çıktı katmanı bulunmaktadır. Giriş

katmanında ham veri analize alınmaktadır. Gizli katmanda giriş değişkenleri aktivasyon fonksiyonuna göre dönüştürülerek işleme alınmakta ve sonuçta çıktı katmanında sınıflandırma yapılmaktadır.

Çok katmanlı perseptronda bağımlı değişken nominal, sıralı, aralık, veya oran ölçeğinde olabilmektedir. Bağımsız değişkenler ise kategorik veya ölçülebilir verilerden oluşabilmektedir. Ölçülebilir veriler şebekenin öğrenmesinin geliştirilmesi maksadıyla yeniden ölçeklendirmeye tabi tutulmaktadır. Frekans ağırlıkları bu süreçte göz ardı edilmektedir.

Çok katmanlı perseptronda kullanılacak veriler $(x-x_{ort})/s$ işleminden geçirilerek standartlaştırılabilmekte, $(x-x_{min})/(x_{maks}-x_{min})$ yapılarak normalize edilebilmekte, $[2*(x-x_{min})/(x_{maks}-x_{min})]-1$ ile uyarlanmış olarak normalize edilmekte veya herhangi bir işlemden geçirmeden kullanılabilmektedir.

Verilerin eğitim, test ve geçerleme örneklem gruplarına bölünmesi yapılarak değişkenlerin rasgele seçilmeleri suretiyle kurulacak modelin sınaması yapılabilmektedir. Yapay sinir ağları şebekesinin öğrenme süreci için eğitim örneklem grubu kullanılmakta, hata düzeltmelerinin yapılması için test örneklemi kullanılmakta ve elde edilen bilgiler ışığında sınıflandırma sonuçlarının test edilmesinde geçerleme örneklemi kullanılmaktadır. Geçerleme örneklemi modelin kurulmasında kullanılmadığından şebekenin bu örneklem yardımıyla performansının da testi yapılmaktadır.

Gizli katmandan elde edilen değerler giriş değişkenlerinin ağırlıklandırması sonucu elde edilen değerlerden oluşmaktadır. Çok katmanlı perseptron bir veya daha fazla gizli katmandan oluşabilmektedir.

Öğrenmede küçük örneklemler için sinaptik ağırlıklara göre küme öğrenme kullanılabileceği gibi büyük örneklemler için anlık öğrenmede kullanılabilmektedir.

Sinaptik ağırlıkların tahmin edilmesinde iki optimizasyon algoritmasından yararlanılmaktadır. Bunlar ölçeklendirilmiş eşlenik gradyan ve gradyan iniş'dir. Ölçeklendirilmiş eşlenik gradyan küme öğrenme tipinde kullanılmaktadır. Gradyan iniş ise anlık öğrenmede kullanılabilmektedir. Gradyan iniş kuralı delta kuralına benzer, hatta aktivasyon fonksiyonunun türevi bağlantı ağırlıklarına uygulanmadan önce, Delta hatasını düzeltmek için kullanılır. Giriş verilerinin güçlü bir modelden çıkarılmadığı uygulamalarda, bu kural özellikle önemlidir (Elmas 2003: 37).

Eşlenik gradyan öğrenme kuralı ise doğrultunun bir önceki doğrultunun doğrusal bileşimi olarak belirtilmesi temeline dayanmaktadır. Uygun β_k değeri için yeni doğrultu

$$p_{k+1} = -\nabla E_{k+1} + \beta_k p_k \qquad (3.7)$$

şeklindedir. Eşlenik gradyan algoritmasının önemli özelliği belli bir adımdaki gradyanın daha önceki doğrultu vektörlerine göre dik olmasıdır. Her yeni doğrultunun gradyanı bir önceki doğrultuya dik olacağından;

$$p_k \nabla E(x_0 + \alpha p_{k+1}) = 0 \qquad (3.8)$$

yazılabilir. Belirtilen denklemlerden;

$$p_k H p_{k+1} = 0 \qquad (3.9)$$

olarak bulunur. Bu şekilde tanımlanan p doğrultu vektörlerine eşlenik doğrultular denir (Veelenturf 1995: 166; Akın 1997: 53; Hertz vd 1991 :126).

3.8.Yapay Sinir Ağlarının Avantaj ve Dezavantajları

Yapay sinir ağlarının kendine has karakteristik özellikleri ağa bazı durumlarda avantaj sağlarken, bazı durumlarda dezavantajlara da dönüşebilmektedir.

Ünlü matematikçi Kolmogorov'un kısaca kendi ismiyle anılan ve ileri beslemeli ağlara uyarlanabilen teoremine göre; girdi katmanında 2n+1 nörona sahip üç katmanlı bir sinir ağı, girdi sinyallerinin uygun bir dönüşümü ile n boyutlu girdi uzayında herhangi bir fonksiyona tam veya kesin olarak yakınsar (n>2 olmak üzere). Teoremde ne ağırlıklar ne de aktivasyon fonksiyonu hakkında bilgi verilmiş sadece böyle ağların var olduğu gösterilmiştir. Teoremin yapay sinir ağlarına avantaj sağlayan yanı, problem ne kadar karışık olsa da keyfi olarak seçilecek sürekli bir fonksiyonun üç katmanlı ileri beslemeli bir sinir ağı ile modellenebilir olmasıdır (Bayru 2007: 22).

Sinir ağlarının yakınsamayı kullanması perseptron modelleri ile regresyon modelleri arasında önemli bir ilişki ortaya koymaktadır (Warren 1994).

Yapay sinir ağlarının bir avantajlı yanı da daha iyi sonuç almak için verilerin manipüle edilebilmesidir. Yapay sinir ağları doğrusal olmayan modellerdir. Doğrusal olmama ve çoklu ilişkinin (multicollinearity) varlığı tahmin konusunda önemli sorunlar yaratabilmektedir. Ancak bu durum sinir ağlarına önemli avantaj sağlar. İlişkisel yapıdan yola çıkarak öğrenmenin etkinliği artırılmış olur.

Yapay sinir ağlarının dezavantajları incelendiğinde ise öncelikle katman sayısı konusu önem kazanmaktadır. Katman sayısının belirlenmesinde bir kıstas bulunmamaktadır. Yüksek sayıda katman kullanılması ağın sınıflandırmada veya tahmindeki hata payının artmasına sebep olabilir.

Ayrıca öğrenme hızı da önemli bir faktördür. Öğrenme hızının ne kadar olması gerektiği önceden tahmin edilemez. Öğrenme hızının belirlenmesi ile ilgili olarak ağın örnekler üzerindeki hatasının belirlenecek bir değerin altına indirilmesi yeterli görülebilmekte ve bu kısıtla eğitimin tamamlanabileceği öngörülmektedir (Özkaynak 2003: 35).

DÖRDÜNCÜ BÖLÜM
4. VERİLERİN DEĞERLENDİRİLMESİ VE ANALİZİ

Bu çalışmada Birleşmiş Milletlerin Kalkınmışlık ölçümleri yaptığı 155 çok gelişmiş ve orta düzeyde gelişmiş ülkelere ait veriler incelenmiştir. İncelenen veriler, Birleşmiş Milletler Kalkınma Programı resmi web sitesinde bulunan Beşeri Kalkınma Endeksi 2008 yılı raporunda ifade edilen verileridir (UNDP Web sitesi). Bu bölümde öncelikle kalkınmışlık ile ilgili genel bilgiler verilecek, verilerle ilgili kısaca bahsedilecek ve bilahare analizde kullanılan yöntemlerle ilgili bilimsel araştırmalar özetlenerek analiz safhasına geçilecektir.

4.1. Beşeri Kalkınma ve Ülkeler için Kalkınmışlık

Ekonomik büyüme ve kalkınma sözcükleri genelde eş anlamlı kullanılmaktadır. Ekonomik büyüme üretim faktörlerinin kişi başına yıldan yıla daha yüksek bir reel gelir sağlayacak şekilde artması olarak tanımlanmaktadır (Ülgener 1974: 409). Kalkınma kavramı ise; salt üretimin ve kişi başına gelirin artırılması demek olmayıp, azgelişmiş bir toplumda sosyo-kültürel yapının da değiştirilmesi, yenileştirilmesidir (Han ve Kaya 1997:2). Yapılan bu açıklamalardan anlaşılabileceği gibi kalkınma sadece azgelişmiş denebilecek ekonomilerle ilgili bir kavram olduğu halde, büyüme süreci, gelişmiş veya kalkınmış ekonomilerle ilgilidir.

Gaspar'a (1995:208) göre kalkınma; maddi refahı artırmaya yönelik potansiyelin gerçekleştirilmesi, insani acıların azaltılması anlamındadır. Başka bir tanıma göre ise kalkınma; beslenme, sağlık, barınma, istihdam, fiziki çevre, sosyo-kültürel çevre, karar mekanizmalarına katılım, insani saygınlık, ait olma duygusu ve benzeri değişkenleri içermelidir (Bhanoji 1991:1452).

İnsani Kalkınma ise Birleşmiş Milletler Kalkınma Programı, 1990 yılı İnsani Kalkınma Raporunda kişilerin tercihlerini geliştirme süreci olarak tanımlanmıştır. İnsani kalkınmayı sağlayan ekonomik, sosyal, politik ve kültürel alanlardaki etkinlikler insani kalkınmanın boyutları olarak kabul edilmektedir. Ancak insan tercihlerini genişletecek üç temel alandaki gelişmeler insani kalkınmanın temel göstergeleri olarak kabul

edilmektedir. Bunlar; gelir, eğitim, sağlık ve beslenme temel göstergeleridir (UNDP 1990:9).

İnsani kalkınmanın öncelikli hedefi; insani tercihleri artırmak ve kalkınmayı daha demokratik ve paylaşıma dayanan bir hale getirmektir. İnsani tercihler; gelire ulaşma, istihdam imkânları, eğitim, sağlık, temiz ve güvenli fiziksel çevre gibi göstergeleri kapsar. Bu göstergeler ayrıca, karar alma mekanizmalarında paylaşıma ve ekonomik ve politik özgürlüklere de vurgu yapılmaktadır (UNDP 1991:7).

1994 İnsani Kalkınma Raporunda insani kalkınma yeniden ve daha geniş olarak tanımlanmıştır. İnsani kalkınmanın temel amacı, şimdiki ve gelecekteki bütün insanların her alandaki potansiyellerini geliştirip kullanabilmesi için uygun ortam ve fırsatların yaratılmasıdır. İnsani kalkınma süreci sadece insanların kapasitelerinin en iyi şekilde geliştirilmesi ile ilgili değildir. Aynı zamanda sağlanan kapasitenin ekonomik, sosyal, siyasal ve kültürel alanlarda da en iyi şekilde kullanılmasını sağlamaya yönelik bir süreçtir (UNDP 1994:13).

Birleşmiş Milletler tarafından 4 farklı kalkınma endeksi 1990 yılından beri kullanılmaktadır. Bunlar; İnsani Kalkınma Endeksi, Cinsiyete Dayalı Kalkınma Endeksi, Cinsiyeti Güçlendirme Endeksi ve İnsani Yoksulluk Endeksidir. Dört endeksin hesabında da farklı hesaplama yöntemleri ve değişkenler kullanılmaktadır.

Tez çalışmasında kullanılan değişkenlerin İnsani Kalkınma Endeksine ait veriler olması sebebi ile bu çalışmada yalnızca İnsani Kalkınma Endeksi ele alınmıştır.

İnsani Kalkınma Endeksi (Human Development Index-HDI) ilk defa 1990 yılında Birleşmiş Milletler Kalkınma Programı tarafından ortaya atılmıştır. Endeks üç temel gösterge ile oluşturulmuştur. ***Gelir***; satın alma gücü paritesine göre kişi başına düşen GSYİH ile ***Yaşam Beklentisi***; doğumda yaşam beklentisi ile ***Eğitim*** ise; yetişkin okuryazar oranı ve okullaşma oranı ile açıklanmaya çalışılmıştır.

İnsani Kalkınma Raporunda yer alan insani gelişim endeks değerine göre ülkeler insani kalkınma boyutunda sınıflandırılmaktadır. Ülkeler ayrıca endeks değerlerine göre de gruplandırılmaktadır (UNDP 1996: 136).

4.2. Analizde Kullanılan Değişkenler ve Analiz Edilen Ülkeler

Yapılan bu tez çalışmasında indekslerin kullanılması yerine doğrudan elde edilen ham veriler kullanılarak Birleşmiş Milletler tarafından yapılan sınıflandırma; diskriminant analizi, lojistik regresyon analizi ve yapay sinir ağları yöntemleriyle yeniden yapılmıştır. Burada İnsani Kalkınma Raporunda belirtilen değişkenlerin azamisinin kullanılmasına özel önem gösterilmiştir. İnsani Kalkınma Raporunda başlangıçta yapılan sınıflandırmaya uygun olarak Çok Gelişmiş ve Orta Düzeyde Gelişmiş ülke sınıflandırmasından yararlanılmıştır. Analizde kullanılan değişkenler Çizelge 4.1'de gösterilmiştir.

Çizelge 4.1. Analizde Kullanılan Değişkenler

	N
Toplam Nüfus(2005) (topnüf)	154
Kentleşme Oranı(2005) (kentoranı)	155
Kadın Parlamenter Oranı (Toplamın Yüzdesi) (kadınpar)	150
Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2007) (Saglık1)	154
Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2007) (Saglık2)	154
Sağlık Harcamaları Kişi Başına (Satın Alma Gücü Paritesine göre$)(2007) (Saglık3)	152
Doğumda Yaşam Beklentisi(2002-2005) (YasamBek)	151
İlköğretime net kayıt oranı (egitim3)	141
1000 kişiye düşen telefon hattı sayısı (2005) (İletisim1)	152
1000 kişiye düşen cep telefonu abonesi sayısı (2005) (İletisim2)	154
1000 kişiye düşen internet kullanıcısı sayısı (2005) (İletisim3)	153
GSYİH (Milyar Dolar) (2005) (GSYİH1)	152
İthal Edilen Mallar ve Hizmetler (GSYİH %'si olarak) (2005) (import)	146
İhraç Edilen Mallar ve Hizmetler (GSYİH %'si olarak) (2005) (export)	147
Elektrik Tüketimi (Kişi Başına Kw-H olarak)(2004) (elektriktuk)	151
Hapiste Bulunan Kişi Sayısı (2007) (Hapis)	155
Geçerli N	120

Çizelge 4.1'de gösterilen değişkenler toplamda 155 ülkeye ait verileri içermektedir. Verileri içeren çizelge EK-1'de bulunmaktadır. Başlangıçta 28 değişkene göre analiz çalışmalarına başlanmış ancak çıkarılan 12 değişkenin kullanılması ile ülke

sayısı 74'e düştüğünden bu değişkenler analiz dışı bırakılarak 120 ülkenin analize alınmasına çalışılmıştır. İnsani Kalkınma Endeksinde özellikle gelir, eğitim, yaşam beklentisi değişkenleri olmazsa olmaz olarak kullanıldığından 16 değişkenden daha fazla ödün verilmemesi gerektiği düşünülmüş ve analizde bu 16 değişkenin kullanımına karar verilmiştir. Değişkenlerle ilgili bilgiler şu şekildedir:

Toplam nüfus değişkeni ile 1 Temmuz itibariyle elde edilen nüfus bilgileri dikkate alınmıştır. Kentleşme oranı değişkeninde şehir ve metropolitenlerde yaşayan ülke nüfusu toplam nüfusun yüzdesi olarak kullanılmıştır. Veriler ülkelerin resmi veri kaynaklarından alınarak Birleşmiş Milletler Kalkınma Programı tarafından derlenmiştir.

Kadın parlamenter oranı ülkelere ait meclis ve senatolarda görevli bayanların sayısı dikkate alınarak hazırlanmıştır. Kamu ve Özel sağlık harcamaları verileri Dünya Bankası verileridir. Kişi Başına düşen Sağlık harcamaları değişkeni ise Dünya Sağlık Örgütü verileridir.

Doğumda yaşam beklentisi bir bebek doğduğunda yaşa bağlı ölüm oranlarına göre kaç yıl yaşayabileceğinin tahmin değeridir. İlköğretime kayıt oranı yaşa bağlı kalınmaksızın okullara kayıt olan öğrenci sayısını temsil etmektedir.

İletişim ile ilgili tüm veriler Dünya Bankası tarafından yayınlanan verilerdir. GSYİH ülkede yaşayan üreticilerin GSMH'den farklı olarak, bir ülke sınırları içerisinde belli bir zaman içinde, üretilen tüm nihai mal ve hizmetlerin para birimi cinsinden değeridir. GSYİH şu formülle hesaplanmaktadır:

GSYİH = tüketim+ yatırım+ devlet harcamaları+ (ihracat-ithalat) (4.1)

İthal ve ihraç edilen mallar ve hizmetler Dünya Bankası verilerinden derlenmiştir. Kişi başına elektrik tüketimi verileri Birleşmiş Milletler tarafından hazırlanan istatistiklerden derlenmiştir. Hapiste bulunan kişi sayısı ise Uluslararası Af Örgütü (Amnesty International) verileridir.

4.3. Literatürde İncelenen Uygulamalar

Analizlerin yapılmasından önce literatürde sınıflandırmaya yönelik olarak yapılan lojistik regresyon analizi, diskriminant analizi ve yapay sinir ağları uygulamaları incelenmiştir. Bu bölümde bu çalışmalardan kısaca bahsedilecektir.

Balcaen ve Ooghe (2005) iş yaşamındaki başarısızlıkların sınıflandırılmasında son 35 yılda kullanılan istatistiksel teknikler ve bu tekniklere ilişkin problemleri yaptıkları çalışmada ele almışlardır. Yaptıkları çalışmada çoklu diskriminant analizi, logit modeller, şartlı olasılık modelleri ve tek değişkenli analiz yöntemlerini karşılaştırmışlardır.

Soeyoshi ve Hwang (2003) veri zarflama analizi ve diskriminant analizi tekniklerini finansal analiz uygulaması ile karşılaştırmıştır.

Rasson ve Bertholet (1999) şirketlerin verilerini kernel yoğunluk tahmini ile uyarlayarak diskriminant analizini uygulamış, *k*-en yakın komşu algoritması ile karşılaştırmasını yapmıştır.

Sueyoshi (2004) diskriminant analizi ile standart tam sayılı programlama modelleri ve iki aşamalı tam sayılı programlama modellerini kullanarak sınıflandırma başarılarını incelemiştir. Japon bankalarından elde ettiği veriler üzerinde de uygulamasını yapmıştır.

Berg (2007) doğrusal diskriminant analizi, genelleştirilmiş doğrusal modeller ve yapay sinir ağlarını kullanarak firmaların iflas tahminlerini yapmaya çalışmıştır.

Çilan vd (2009) Avrupa Birliği üyesi olan ve olmayanlar arasındaki dijital ayrımın analizini diskriminant analizi ile yapmışlarıdır. Yaptıkları analizde sınıflandırma başarısı %74,1 ile başarılı bulunmuştur. Analiz öncesi normallik varsayımının testi yapılmıştır.

Bosse (2008) çoklu diskriminant analizi ile küçük firmaların borç alırken kredibilitesinin ayrıştırılmasını modellemiş ve %86,6'lık bir sınıflandırma başarısı elde etmiştir.

Wu vd (2008) Çin kamu şirketlerinin finansal olumsuzluklarının analizini olasılıklı yapay sinir ağları ve diskriminant analizi kullanarak yapmıştır. Kısa dönem tahminlerde çoklu diskriminant analizi ile %81,25, uzun dönem tahminlerde %56,25'lik bir sınıflandırma başarısı elde etmiştir. Buna karşılık yapay sinir ağları ile yapılan analiz sonucunda kısa dönemde %87,5'lik, uzun dönemde %81,25'lik bir sınıflandırma başarısı elde etmişlerdir.

Chen vd (2008) iletişim aracının seçimi ile ilgili olarak belirlenen kriterleri diskriminant analizi ile analiz etmiş ve tahmin modeli geliştirerek değişkenler arasındaki ilişkiyi incelemiştir.

Pompe ve Bilderbeek (2005) küçük ve orta ölçekli sanayi firmalarının iflasını tahmin etmede çoklu diskriminant analizi ile geliştirdiği diskriminant modelini kullanmıştır.

Pasiouras ve Tanna (2009) bankaların tedarik hedeflerinin tahmini için lojistik regresyon (Logit model) ve diskriminant analizi tekniklerini kullanmışlardır. Altı farklı araştırma modeli analize tabi tutulmuş ve sonuçları karşılaştırılmıştır.

Allen vd (2005) adımsal diskriminant analizini ülkeler arası ekonomik inançlardaki farkların tespit edilmesinde kullanmıştır.

Emel vd (2003) ticari banka sektöründe kredi derecelendirmesi için bir yaklaşım olarak veri zarflama analizi ve diskriminant analizinden yararlanmışlardır.

Pollalis (2003) bilgi yoğun firmalarda karşılıklı birlikteliği birleşme stratejileri ışığında ortaya koymaya çalışmıştır. Yaptığı analizde kümeleme analizi ve diskriminant analizini kullanmıştır.

Malhotra ve Malhotra (2003) müşteri borçlarının değerlendirilmesinde çoklu diskriminant analizi ve yapay sinir ağlarını kullanmıştır. Uygulamada kullanılan 5 örneklemde yapay sinir ağlarının sınıflandırmada daha başarılı olduğu görülmüştür.

Cheng ve Titterington (1994) farklı yapay sinir ağları modelleri ile istatistiksel metotları karşılaştırmıştır. İleri beslemeli ağlar ile diskriminant analizi ve lojistik regresyon arasında güçlü bir ilişki olduğunu göstermişlerdir.

Odom ve Sharda (1990) yapay sinir ağları ve çok değişkenli diskriminant analizinin tahmin kabiliyetini karşılaştırmışlardır. Yapay sinir ağları azalan örneklem üzerine yapılan analiz yönteminde daha iyi performans gösterdiği tespit edilmiştir.

Lesho ve Spector (1996) yapay sinir ağlarının tahmin yeteneğini doğrusal diskriminant analizi ve karesel diskriminant analizi ile karşılaştırmıştır. Seçilen yapay sinir ağları modellerinin klasik diskriminant analiz modellerinden daha doğru sonuçlar verdiği gözlenmiştir.

Lee vd (2005) Kore firmalarının iflasının tahmin edilmesinde yapay sinir ağlarını kullanmışlardır. Diskriminant analizi ve lojistik regresyon analizi tahmin doğruluğu açısından karşılaştırılmıştır. Geriye yayılım algoritması ile kullanılan ağ şayet hedef vektörü belirli ise örneklem sayısı küçülse bile en iyi çözümü vermektedir.

Limsonbunchai vd (2005) lojistik regresyon ve yapay sinir ağları modellerini Tayland'daki tarım borçlarının kredi derecelendirmesinde kullanmıştır. Kredi

derecelendirmede olasılıklı yapay sinir ağlarından yararlanmışlardır. Logit model ve çok katmanlı ileri beslemeli yapay sinir ağı modeli kullanılmıştır. Yapay sinir ağları bu çalışmada lojistik regresyon modeli sonuçlarına göre daha kötü sonuçlar vermiştir.

Ainscough ve Aronson (1999) yapay sinir ağları ile regresyon analizini modelleme ve tahmin açısından incelemiştir. Çalışma sonuçlarına bakıldığında yapay sinir ağlarının daha etkili olduğu ve tahmin sonuçlarının daha iyi performans gösterdiği gözlenmiştir.

Gan vd (2005) olasılıklı yapay sinir ağları ve çok katmanlı ileri beslemeli yapay sinir ağları ile elektronik bankacılık kullanan ve kullanmayan müşteriler arasındaki tercihlerin lojistik modelini karşılaştırmışlardır. Deneysel sonuçlara göre alıkoyma örneklemi için çok katmanlı ileri beslemeli yapay sinir ağları ile lojistik modelin hemen hemen eşit seviyede tahmin performansı gösterdiği tespit edilmiştir. Olasılıklı yapay sinir ağları bu çalışmada en iyi sonucu veren model olarak gösterilmiştir.

Literatürdeki çalışmalar da kısaca özetlendikten sonra bahse konu veriler kullanılarak yapılan analizler ve sonuçları bundan sonraki bölümden itibaren izah edilecektir.

4.4. Uygulamanın Konusu ve Amacı

Bu çalışmada uygulama olarak Birleşmiş Milletler Kalkınma Programı tarafından beşeri kalkınma indeksinde çok gelişmiş ve orta düzeyde gelişmiş toplam 155 ülkenin istatistiksel yöntemlerden diskriminant analizi ve lojistik regresyon analizi ile yeniden sınıflandırılması yapılmıştır. Ayrıca yapay sinir ağları ile modelleme yapılmış ve sınıflandırma yeniden değerlendirilmiştir.

Bu çalışmanın amacı istatistiksel metotlar ile yapay sinir ağlarının sınıflandırma başarısının incelenmesidir. Ayrıca 155 ülkeye ait 16 değişkenin sınıflandırmadaki önem derecelerinin tespiti de yapılmıştır.

Yapılan çalışma Birleşmiş Milletlerin uyguladığı yönteme bir alternatif olarak değerlendirilmektedir. Yapılan farklı analiz yöntemleri ile elde edilen sınıflandırma başarılarının yorumlanması sonucunda literatüre katkı yapmak hedeflenmektedir. Orta düzeyde gelişmiş olan ülkelerin de belirlenen 16 değişkenden hangisine daha fazla önem vererek gelişmiş ülkeler sınıfına dâhil edilebileceği de ayrıca vurgulanmaktadır.

4.5. Diskriminant Analizi Uygulama Sonuçlarının Analiz Edilmesi

155 ülkenin değerleri kullanılarak diskriminant analizi yapılmış ve sonuçlar elde edilmiştir. Değişken değerleri dikkate alındığında toplam 155 değişkenin 35'inin işleme dâhil edilmediği ve 120 değişken ile sınıflandırma sürecinin yürütüldüğü Çizelge 4.2'de gösterilmektedir. Analizde bağımlı değişken olarak çok gelişmiş ve orta düzeyde gelişmiş ülke sınıflandırması kullanılmıştır. Ayrıca bağımsız değişken olarak ta 16 bağımsız değişken analize alınmıştır.

Çizelge 4.2. İşleme Alınan Örneklem Sayısı

Ağırlıklandırılmamış Durumlar		N	Yüzde
Geçerli		120	77,4
Çıkarılan	Kayıp veya aralık dışı kodlamalar	0	0
	En az bir kayıp ayırt edici değişken	35	22,6
	Kayıp veya aralık dışı kodlamalar(İkisi Birlikte)	0	0
	Toplam	35	22,6
Toplam		155	100,0

Ayrıca ülkelerin önceden belirlenmiş gruplara göre olasılıkları da Çizelge 4.3'de görülmektedir. Örneklem büyüklüğü 120, çok gelişmiş grup içerisinde 56, orta düzeyde gelişmiş grup içerisinde de 64 ülke bulunmaktadır.

Çizelge 4.3. Grupların Öncelik Olasılıkları

Ülkelerin Gelişmişlik Sınıflandırması	Öncelikli	Analizde Kullanılan Durumlar	
		Ağırlıklandırılmamış	Ağırlıklandırılmış
Çok Gelişmiş	0,467	56	56,000
Orta Düzeyde Gelişmiş	0,533	64	64,000
Toplam	1,000	120	120,000

Grupların öncelikli olasılıkları sırasıyla Çok Gelişmiş sınıfı için 0,467 ve Orta Düzeyde Gelişmiş sınıfı için 0,533 olarak tespit edilmiştir. Bu olasılıklar daha sonra sınıflandırma oranının değerlendirilmesinde kullanılacaktır. Diskriminant analizi ile ilgili olarak analiz sonuçlarına başlanmadan önce normallik varsayımı, kovaryans matrislerinin eşitliği varsayımı ve çoklu bağlantı varsayımı incelenecek, daha sonra sınıflandırma analizi ele alınacaktır.

4.5.1. Diskriminant analizi varsayımları

Diskriminant analizi ile ilgili literatürde de bahsedildiği gibi çok önemli üç temel varsayım analiz öncesi araştırılmakta, elde edilen değerlere göre analiz yapılmamakta veya farklı yöntemler kullanılarak analize devam edilmektedir. Bu varsayımların başında çok değişkenli normallik, kovaryans matrislerinin eşitliği ve çoklu bağlantı varsayımı gelmektedir. Varsayımların sağlanamamasının elde edilecek sınıflandırma sonuçları açısından sorun yaratacağı ve arzu edilen yüksek oranlarda sınıflandırma yapılamayacağı literatürde ifade edilmektedir.

Kovaryans matrislerinin eşit olması durumunda doğrusal diskriminant analizi yapılabilirken kovaryans matrislerinin eşit olmaması durumunda kuadratik diskriminant analizi yapılarak sınıflandırma sonuçları elde edilebilmektedir.

4.5.1.1. Normallik varsayımı

Çok değişkenli normalliğin tespit edilmesinde Sharma (1996:380-382) ana kütleler normal ve örneklem büyüklüğü (n>25) ise bu durumda Mahalanobis uzaklıklarının Ki-kare dağılımına uyduğunu ifade etmektedir. Başlangıçta değişkenlerin normal dağılıp dağılmadığı araştırılmış ve bazı değişkenlerin normal dağılmadığı tespit edilmiştir. Bahse konu değişkenlerin normalleştirilmesi maksadıyla değişken değerlerinin logaritması alınmış ve normalize oldukları görülmüştür. Kolmogorov-Smirnov testi sonucunda normal dağlım göstermeyen değişkenler şunlardır:

- Toplam Nüfus (Topnüflog)
- Satın Alma Gücü Paritesine Göre Kişi Başına Sağlık Harcamaları (Saglık3log)
- GSYİH (Milyar Dolar) (GSYİH1log)
- İhraç Edilen Mallar ve Hizmetler (GSYİH'nın %'si olarak) (ihraçlog)
- İthal Edilen Mallar ve Hizmetler (GSYİH'nın %'si olarak) (ithallog)
- Hapiste Bulunan Kişi Sayısı (Hapislog)
- Kişi Başına Elektrik Tüketimi (Elektriklog)
- 1000 kişiye düşen telefon hattı sayısı (2005) (İletisim1log)

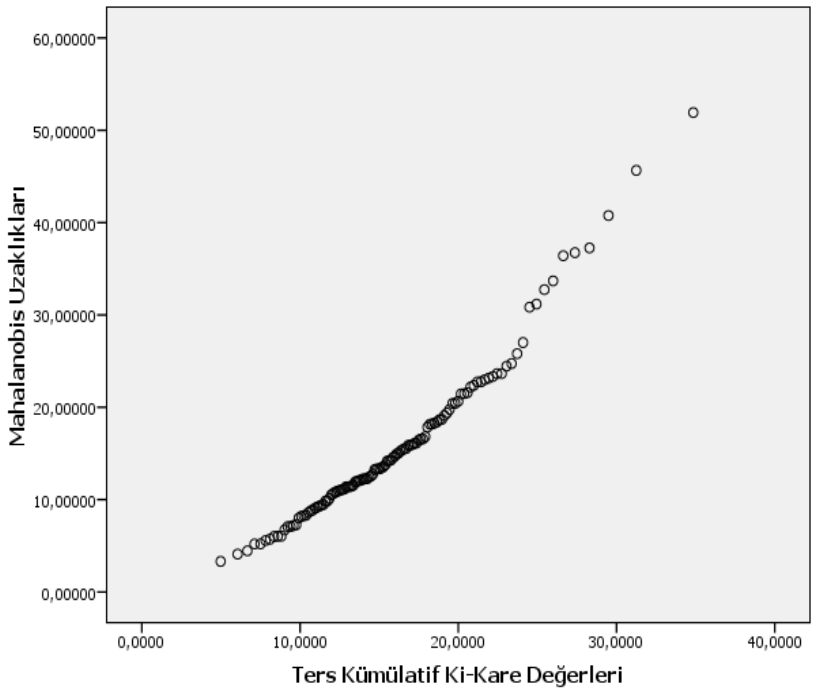
Belirtilen değişkenlerin logaritması alınarak normallik dönüşümü yapılmıştır. Tekrarlanan testte değişkenlerin tamamının tek değişkenli normallik gösterdiği

görülmüştür. Dönüştürülen değişkenler yukarıda değişkenlerin karşılarında belirtilen yeni kodlarla tanımlanmıştır.

Normalleştirilen verilerin kullanılması ile Sharma (1996)'nın belirttiği çok değişkenli normallik analizi sonucu Çizelge 4.4'te görülmektedir. Mahalanobis uzaklıkları ile ters kümülatif ki-kare değerleri arasında 0,979'luk bir korelasyon vardır ve 0,01 anlamlılık düzeyinde bu değer anlamlıdır. Aynı durum Şekil 4.1'de serpilme diyagramında da görülebilmektedir.

Çizelge 4.4. Mahalanobis Uzaklıkları ve Ki-Kare Değerleri Korelasyon Analizi Çizelgesi

		Mahalanobis Uzaklıkları	Ters Kümülatif Ki-Kare Değerleri
Mahalanobis Uzaklıkları	**Pearson Korelâsyon**	**1,000**	**0,979****
	Anl. (2-taraflı)		**0,000**
	N	**120,000**	**120**
Ters Kümülatif Ki-Kare Değerleri	**Pearson Korelâsyon**	**0,979****	**1,000**
	Anl. (2-taraflı)	**0,000**	
	N	**120**	**120,000**
****. 0.01 düzeyinde Korelasyon anlamlıdır 0.01**			



Şekil 4.1. Mahalanobis Uzaklıkları ve Ki-Kare Değerleri Korelasyon Diyagramı

Mahalanobis uzaklıkları ile ters kümülatif ki-kare değerleri arasındaki korelasyon değerinin 1'e eşit olması arzu edilmektedir. Elde edilen korelasyon değeri 1'e yakın bir değer olduğundan bu aşamada çok değişkenli normal dağılım varsayımının sağlandığı düşünülmektedir. Ayrıca korelâsyon katsayısının normallik testi için olasılık tablosunda n=100 için 0,01 olasılık düzeyindeki tablo değeri 0,981'dir. Elde edilen korelasyon katsayısı bu değere oldukça yakındır.

Çok değişkenli sapan birim değerlerinin incelenmesi ile de çok değişkenli normallik testinin yapılabildiği Kalaycı (2008: 212-214) tarafından ifade edilmektedir. Sapan birimlerin incelenmesinde de yine Mahalanobis uzaklıkları hesaplanmakta analizde kullanılan değişken sayısına hesaplanan uzaklıklar bölünerek sapma değerleri bulunmaktadır. Bulunan bu değerlerin t dağılımına uyduğu belirtilmektedir. Herhangi bir birimin sapan değer olarak belirlenebilmesi için ilgili birimin %1 anlamlılık seviyesinde anlamlı olması gerekmektedir. Yani MD^2/sd değerinin (t) 5,014'den büyük olması gerekmektedir. Yapılan incelemede analizde kullanılan 120 birimin hiçbirisi sapan değer olarak değerlendirilmemiştir.

4.5.1.2. Kovaryans matrislerinin eşitliği varsayımı

Kovaryans matrislerinin eşitliği için Box M testi kullanılmıştır. Yapılan analizde kovaryans matrislerinin eşitliği sağlanamamıştır ($p<0,05$). Bu sebeple kovaryans matrislerinin eşit olmadığı varsayımından yola çıkılarak Kuadratik Diskriminant Analizi kullanılmıştır. Box M testi sonucu Çizelge 4.5'de bulunmaktadır.

Çizelge 4.5 Box M Test Sonuçları

Box M		**440,561**
F	**Yaklaşık**	**2,775**
	sd1	**136**
	sd2	**41,455.39425642**
	Anl.	**1.004113554704564E-20**

Kovaryans matrislerinin eşitliği gruplar arasında sağlanamadığından kuadratik formülasyon dikkate alınmıştır.

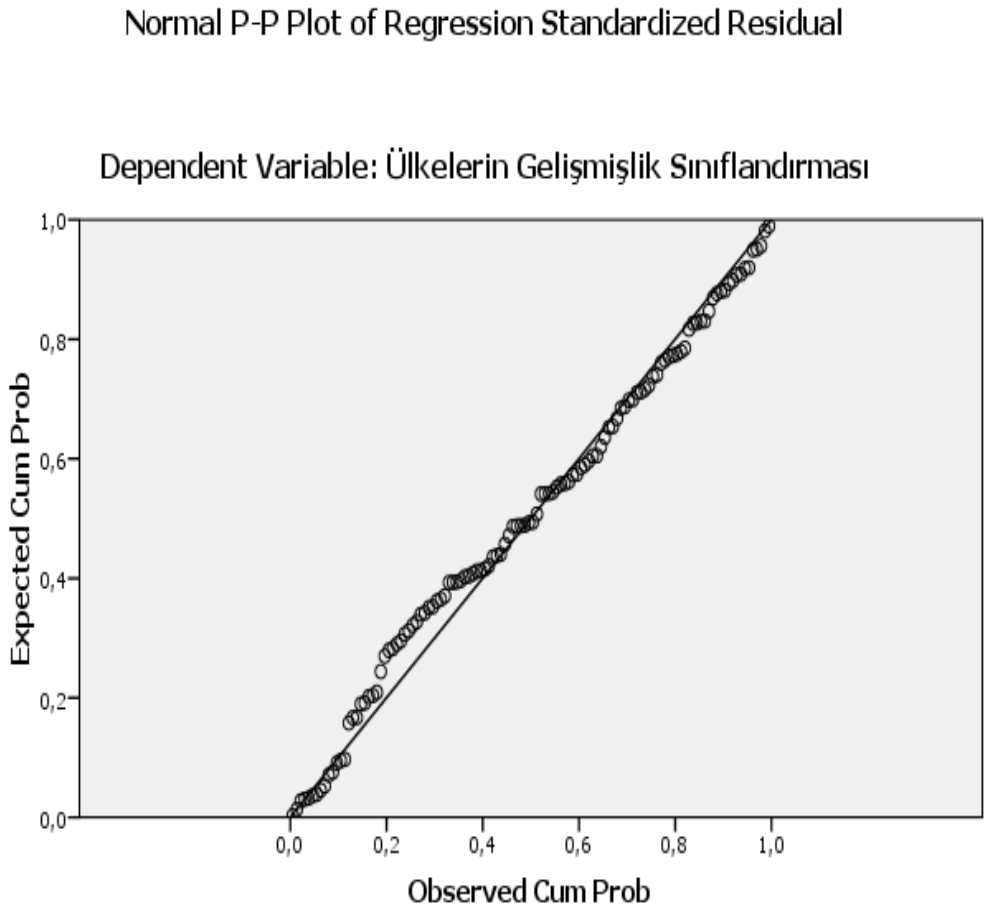
4.5.1.3. Çoklu bağlantı varsayımı

Çoklu bağlantının belirlenebilmesi maksadıyla doğrusal regresyon analizi ile doğrusallık (çoklu İlişki) araştırılmıştır. Bu analizden amaç şayet değişkenlerin bazıları birbirleriyle yüksek korelâsyona sahip ise tahmin gücünün azaltılmaması için ya bu değişkenlerden birisinin analiz dışı bırakılması ya da birlikte kullanılabilme durumu var ise tek bir değişkene dönüştürülerek kullanılmasına karar vermektir. Aksi takdirde katsayılar tanımsız ve bu katsayıların standart hatası sonsuz olabilecektir. Ayrıca katsayıların varyans ve kovaryansları da artma eğilimi gösterecektir. Bu durumların oluşmaması için analiz yapılarak verilerin incelenmesi önem kazanmaktadır. Araştırma sonuçları Çizelge 4.6'da bulunmaktadır.

Çizelge 4.6. Çoklu Doğrusallık Testi

Model		Standartlaştırılmamış Katsayılar		Standartlaştırılmış Katsayılar	t	Anl.	Çoklu İlişki İstatistiği	
		B	St.Hata	Beta			Tolerans	VIF
1	(Sabit)	2,627	0,392		6,705	0,000		
	Toplam Nüfus(2005)	3,104E-5	0,000	0,007	0,114	0,909	0,750	1,334
	Kentleşme Oranı(2005)	0,001	0,002	0,029	0,360	0,719	0,411	2,434
	Kadın Parlamenter Oranı (Toplamın Yüzdesi)	0,000	0,003	-0,009	-0,132	0,895	0,645	1,550
	Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2004)	-0,028	0,026	-0,119	-1,090	0,278	0,228	4,383
	Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2004)	-0,006	0,028	-0,018	-0,226	0,822	0,447	2,236
	Sağlık Harcamaları Kişi Başına (Satın Alma Gücü Paritesine göre US$)	0,000	0,000	0,285	1,787	0,077	0,108	9,276
	Doğumda Yaşam Beklentisi	-0,011	0,007	-0,169	-1,677	0,097	0,270	3,710
	İlköğretime net kayıt oranı	0,003	0,003	0,064	0,872	0,385	0,510	1,962
	1000 kişiye düşen telefon hattı sayısı	0,000	0,000	-0,355	-2,543	0,012	0,141	7,104
	1000 kişiye düşen cep telefonu abonesi sayısı (2005)	0,000	0,000	-0,393	-3,557	0,001	0,225	4,449
	1000 kişiye düşen internet kullanıcısı sayısı (2005)	0,000	0,000	-0,089	-0,734	0,465	0,184	5,421
	GSYİH (Milyar Dolar) (2005)	7,107E-6	0,000	0,018	0,140	0,889	0,169	5,921
	İthal Edilen Mallar ve Hizmetler (GSYİH %'si olarak) (2005)	0,005	0,002	0,227	2,020	0,046	0,218	4,597
	İhraç Edilen Mallar ve Hizmetler(GSYİH %'si olarak) (2005)	-0,005	0,003	-0,230	-1,913	0,058	0,190	5,256
	Kişi Başına Elektrik Tüketimi	-6,240E-6	0,000	-0,060	-0,618	0,538	0,288	3,468
	Hapiste Bulunan Kişi Sayısı (2007)	-2,480E-7	0,000	-0,108	-0,937	0,351	0,207	4,838

Çoklu doğrusallık testinde VIF ve Tolerans değerlerinin incelenmesinde VIF değerlerinin 10'dan küçük olduğu ve Tolerans değerlerinin 0,30'a yakın veya üzerinde olduğu gözlenmektedir. Bu durum çoklu doğrusal ilişkinin olmadığı yönünde yorumlanabilmektedir. Ayrıca t değerlerinin çok küçük değer almasının da çoklu doğrusallık sorununa işaret ettiği bazı yazarlarca ifade edilmektedir. Çizelge incelendiğinde t değerlerinden 0'a çok yakın değerler bulunmadığı da ayrıca gözlenebilmektedir.



Şekil 4.2. Ülkelerin Gelişmişlik Sınıflandırması

Şekil 4.2, doğrusal çizginin altındaki ve üstündeki noktaların değişkenin normal dağılımdan hangi düzeyde bir sapma gösterdiğini görselleştirmektedir. Şekil 4.2'de görüldüğü gibi ülkelerin gelişmişlik sınıflandırması değişkenine ait regresyon artık değerleri normallikten önemli bir sapma göstermemektedir.

4.5.2. Diskriminant fonksiyonlarının önem derecesinin tespiti

Başlangıçta belirlenen iki grup (Çok Gelişmiş ve Orta Düzeyde Gelişmiş) olduğu için 1 Diskriminant fonksiyonu türetilmiştir. Özdeğerin (Eigenvalue) büyük olması bağımlı değişkendeki varyansın daha büyük bir kısmının elde edilen fonksiyon tarafından açıklanabildiğini göstermektedir. Kesin bir değer olmamakla birlikte 0,40'ın üzerindeki değerler iyi olarak kabul edilmektedir. Çizelge 4.7'de görülebileceği gibi modelde özdeğer 2,385 bulunmuş ve varyansın %100'ünü açıklamaktadır. Ayrıca Kanonik Korelasyon Katsayısı 0,839 olarak bulunmuştur. Katsayının karesi 0,704'dür. Bağımsız değişkenlerin bağımlı değişkeni %70,4 oranında açıkladığı söylenebilir.

Çizelge 4.7. Özdeğerler Çizelgesi

Fonksiyon	Özdeğer	Varyansın Yüzdesi	Kümülatif %	Kanonik Korelasyon
1	2,385	100,0	100,0	0,839

Çizelge 4.8. Wilk's Lambda Değeri

Fonksiyon testi	Wilks' Lambda	Ki-Kare	Sd	Anl.
1	0,295	134,142	16	1.00411355470E-20

Wilk's Lambda istatistiği, diskriminant skorlarındaki toplam varyansın gruplar arasındaki farklar tarafından açıklanamayan kısmını (oranını) göstermektedir. Modelde 0,295 yani toplam varyansın %29,5'u gruplar arasındaki farklar tarafından açıklanamamaktadır.

4.5.3. Diskriminant fonksiyonu ve yorumu

Elde edilen bir adet fonksiyon ve fonksiyon içerisinde bulunan değişkenlerin katsayıları Çizelge 4.9'da bulunmaktadır.

Çizelge 4.9. Standartlaştırılmamış Diskriminant Fonksiyonu Katsayıları

	Fonksiyon
Topnüflog	**0,922**
saglik3log	**2,676**
iletisim1log	**-0,528**
iletisim3log	**0,160**
GSYİH1log	**-0,495**
İthallog	**-0,565**
İhraçlog	**0,385**
Elektriklog	**0,440**
Hapislog	**-0,424**
Kentleşme Oranı(2005)	**-0,002**
Kadın Parlamenter Oranı (Toplamın Yüzdesi)	**0,001**
Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2004)	**-0,127**
Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2004)	**-0,189**
Doğumda Yaşam Beklentisi(2002-2005)	**0,045**
İlköğretime net kayıt oranı	**-0,019**
1000 kişiye düşen cep telefonu abonesi sayısı (2005)	**0,002**
(Sabit)	**-7,119**

Ülkelerin gelişmişlik düzeylerinin belirlenmesinde kullanılan diskriminant fonksiyonu Z gelişmişlik düzeyini belirlemek üzere (1 veya 2) şu şekilde oluşturulmuştur:

Z= -7,119+**0,922***TopNüflog+ **2,676***Sağlık3log (Sağlık Harcamaları Kişi Başına (Satın Alma Gücü Paritesine göre US$)(2004)) **-0,002***Kentleşme Oranı(2005)+ **0,01***Kadın Parlamenter Oranı (Toplamın Yüzdesi)**-0,127***Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)**-0,189***(2004 Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2004)+**-0,045***Doğumda Yaşam Beklentisi(2002-2005)**-0,019***İlköğretime net kayıt oranı-**0,528***iletisim1log(1000 kişiye düşen telefon hattı sayısı (2005)) **+0,002***1000 kişiye düşen cep telefonu abonesi sayısı (2005)+**0,160***iletisim3log(1000 kişiye düşen internet kullanıcısı sayısı (2005))-**0,495***GSYİH1log(GSYİH (Milyar Dolar) (2005))-**0,565***İthallog(İthal Edilen Mallar ve Hizmetler (GSYİH %'si olarak) (2005))+**0,385***İhraçlog(İhraç Edilen Mallar ve Hizmetler (GSYİH %'si olarak) (2005))+**0,440***Elektriklog(Kişi Başına Elektrik Tüketimi (Kw-H olarak)(2004)) – **0,424***Hapislog (Hapiste Bulunan Kişi Sayısı (2007))

Modelden görülebileceği gibi 1 birimlik artış ile bağımlı değişken üzerinde en büyük etki yaratan değişken Kişi Başına Satın Alma Gücü Paritesine göre Sağlık

Harcamaları (Sağlık3log)'dır. 1 birimlik artışla 2.676'lık bir pozitif etki yaratmaktadır. İthalatın yüksek oluşunun negatif etkisi olduğu ihracatın ise pozitif etki yarattığı görülmektedir. Ayrıca elektrik tüketiminin ve iletişim değişkenlerinin de pozitif etki yarattığı söylenebilir.

Ayrıca sınıflandırma sonuçlarına göre elde edilen diskriminant fonksiyonu ile hangi ülkenin doğru sınıflandırılmaya tabi tutulduğu ve hangilerinin yanlış sınıflandırıldığı da Z skorlarının incelenmesi ile görülebilecektir.

Her bir grubun ortalama diskriminant fonksiyonu ise Çizelge 4.10'da görülmektedir. Buna göre birinci grubun ortalama değeri ve ikinci grubun ortalama değerlerinin fonksiyona uzaklıkları sırasıyla 1,637 ve –1,433'tür.

Çizelge 4.10. Grup Ortalamalarının Diskriminant Fonksiyonuna Olan Uzaklıkları

	Fonksiyon
Ülkelerin Gelişmişlik Sınıflandırması	**1**
Çok Gelişmiş	**1,637**
Orta Düzeyde Gelişmiş	**-1,433**

4.5.4. Diskriminant analizinde bağımsız değişkenlerin öneminin değerlendirilmesi

Diskriminant analizinde diskriminant fonksiyonu ortaya konduktan sonra fonksiyon üzerindeki etkileri yapı matrisi ile incelenmektedir. Yapı Matrisi bağımsız değişkenlerin öneminin değerlendirilmesinde kullanılan bir matristir. Yapı matrisi her bir değişkenin diskriminant fonksiyonu ile olan korelâsyonunu göstermektedir.

Çizelge 4.11. Yapı Matrisi

	Fonksiyon
1000 kişiye düşen cep telefonu abonesi sayısı (2005)	**0,838**
saglik3log	**0,817**
Elektriklog	**0,689**
iletisim1log	**0,637**
iletisim3log	**0,637**
Doğumda Yaşam Beklentisi(2002-2005)	**0,598**
Kentleşme Oranı(2005)	**0,427**
Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2004)	**0,421**
GSYİH1log	**0,414**
İlköğretime net kayıt oranı	**0,276**
Kadın Parlamenter Oranı (Toplamın Yüzdesi)	**0,248**
İhraçlog	**0,147**
Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2004)	**-0,123**
Hapislog	**0,057**
İthallog	**-0,045**
Topnüflog	**0,011**

Çizelge 4.11'de bulunan yapı matris değerleri incelendiğinde Modelde bulunan değişkenlerden 1000 kişiye düşen cep telefonu aboneliği sayısı, Satın Alma Gücü Paritesine Göre Kişi Başına Sağlık Harcamaları (Sağlık3log), Kişi Başına Elektrik Tüketimi (elektriklog), Doğumda Yaşam Beklentisi, 1000 Kişiye Düşen Telefon Hattı Sayısı (iletisim1log), 1000 Kişiye Düşen İnternet Hattı(iletisim3log) değişkenlerinin modeldeki fonksiyonla etkili bir korelâsyona sahip olduğu görülebilmektedir. Değişkenler kategorik olarak değerlendirildiğinde sağlık, enerji ve iletişim değişkenlerinin sınıflandırma ve ayırt etmede daha etkili olduğu izlenimi vermektedir. Ekonomi değişkenlerinin bu sıralamada daha gerilerde olmasının önemli olarak değerlendirilebileceği düşünülmektedir.

4.5.5. Sınıflandırma sonuçları

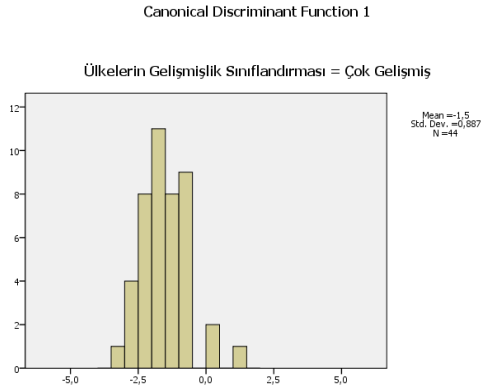
Çizelge 4.12. Sınıflandırma Sonuçları

			Tahmin Edilen Grup Üyeliği		
		Ülkelerin Gelişmişlik Sınıflandırması	Çok Gelişmiş	Orta Düzeyde Gelişmiş	Toplam
Orjinal	Sayı	Çok Gelişmiş	51	5	56
		Orta Düzeyde Gelişmiş	4	60	64
	%	Çok Gelişmiş	91,1	8,9	100,0
		Orta Düzeyde Gelişmiş	6,2	93,8	100,0

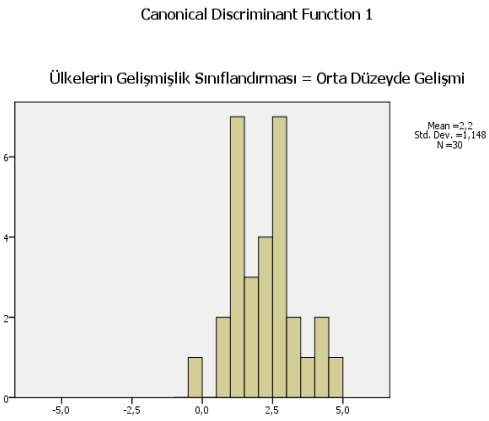
Oluşturulan modelin %92,5'lik toplam sınıflandırma oranı ile başarılı bir sınıflandırma yaptığı söylenebilir. Ancak bu sınıflandırmanın doğruluğunun test edilmesi maksadıyla nisbi şans kriteri ve maksimum şans kriterinin hesaplanarak karşılaştırılması gerekmektedir. Hesaplamaya alınan örneklem büyüklüğü 120'dir. Dolayısıyla Çok Gelişmiş grup örneklemin %47'sini, orta gelişmiş grup ise %53'ünüi oluşturmaktadır. Şans değeri çok gelişmiş grubun seçilme ihtimali yani 0,47 ve orta gelişmiş grubun seçilme ihtimali yani 0,53'tür. Burada maksimum şans kriteri 0,53'tür. Nisbi şans kriteri ise $(0,47)^2+(0,53)^2=0,5018$'dir. Diskriminant analizi sonucunda elde edilen sınıflandırma oranı bu değerlerin çok üzerindedir.

Sınıflandırmada hatalı sınıflandırılan 9 ülke bulunmaktadır. Çok gelişmiş ülke grubunda iken elde edilen diskriminant modeli ile orta gelişmiş düzeyde olarak 5 ülke, Uruguay, Meksika, Panama, Beyaz Rusya ve Arnavutluk bulunmuştur. Orta gelişmiş

ülkeler arasında olup ta diskriminant analizinin çok gelişmiş ülke grubuna ise 4 ülkeyi, Türkiye, Kolombiya, Tunus ve Jamaika'yı atadığı gözlenmektedir. Değişkenlerin sıra numaraları aynı zamanda Birleşmiş Milletler Kalkınma Programında belirlenen gelişmişliğe bağlı sıra numaralarıdır. Bu durumda hatalı sınıflandırılan ülkelerin çok gelişmiş ülkeler ile orta düzeyde gelişmiş ülkelerin sınır noktasına yakın olmasının yapılan analizin Birleşmiş Milletler Kalkınma Programınca yapılan sınıflandırmaya çok yakın bir çalışma olduğunu göstermektedir. Analizde Türkiye'nin belirlenen bağımsız değişken değerlerine göre Çok Gelişmiş olarak sınıflandırılırken Birleşmiş Milletler Kalkınma Programı tarafından Orta Gelişmişlik düzeyinde sınıflandırılmasının da anlamlı bir sonuç olabileceği düşünülmektedir.



Şekil 4.3. Çok Gelişmiş Ülke Grubunun Dağılım Grafiği



Şekil 4.4. Orta Düzeyde Gelişmiş Ülke Grubunun Dağılım Grafiği

Şekil 4.3 ve Şekil 4.4'de de grupların kanonik diskriminant fonksiyonuna göre dağılımları görülebilmektedir.

4.5.6. Adımsal diskriminant analizi sonuçları

Adımsal diskriminant analizi hangi değişkenlerin ayırt etmede daha etkili olduğunun belirlenmesi için yapılmıştır. Bu sebeple kuadratik diskriminant analizi yapılırken olduğu gibi detaylı bir şekilde konunun gösterilmesi yerine kullanılan yönteme göre hangi değişkenlerin modelde kaldığı ve modelde kalan değişkenlerin sınıflandırma başarısının ne olduğu üzerinde durulacaktır.

Adımsal diskriminant analizinde değişken seçiminde Wilks metodu, Mahalanobis metodu ve En küçük F oranı metotları kullanılmıştır. Bu metotlara göre modele dâhil edilen değişkenler ve model katsayıları Çizelge 4.13'te gösterilmiştir.

Çizelge 4.13. Adımsal Diskriminant Analizi Karşılaştırma Çizelgesi

	Mahalanobis Metodu		En Küçük F Oranı Metodu		Wilks Metodu	
	Fonksiyon	Model Değerleri	Fonksiyon	Model Değerleri	Fonksiyon	Model Değerleri
Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2004)	-0,158	Özdeğer= 2.030 r=0,821 Varyansın Yüzdesi=%100 Wilks Lamda=0,326 Ki-Kare=131,244 Sd=2 p<0,05	-0,158	Özdeğer= 2.030 r=0,821 Varyansın Yüzdesi=%100 Wilks Lamda=0,326 Ki-Kare=131,244 Sd=2 p<0,05 F=121.077 sd1=2 sd2=117 p<0,05	-0,158	Özdeğer= 2.030 r=0,821 Varyansın Yüzdesi=%100 Wilks Lamda=0,326 Ki-Kare=131,244 Sd=2 p<0,05
Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2004)	-0,238		-0,238		-0,238	
saglik3log	2,744		2,744		2,744	
1000 kişiye düşen cep telefonu abonesi sayısı (2005)	0,001		0,001		0,001	
(Sabit)	-6,641		-6,641		-6,641	

Çizelge 4.13'de görülebileceği gibi modelde 4 değişkenin bulunması ile grupların sınıflandırılabileceği belirtilmektedir. Genel olarak model incelendiğinde iletişim ve sağlık değişkenlerinin çok önemli bir yer tuttuğu görülebilir. Ancak Kamu ve Özel sağlık harcamalarının modelde negatif etkisi olurken kişi başına satın alma gücü paritesine göre sağlık harcamalarının (saglik3log) çok önemli hatta yegâne belirleyici olduğu bile söylenebilir.

Kurulan bu modele göre sınıflandırma başarıları izlendiğinde ise tüm metotların %90,7'lik bir sınıflandırma başarısı yakaladığı gözlenmiştir. Bu durumun da dikkat çekici olduğu düşünülmektedir. Durumlara ilişkin değerler incelendiğinde ise 3 metot ile yapılan adımsal analizde de Türkiye'nin modeldeki değişken değerlerine göre Çok Gelişmiş ülkeler arasında sınıflandırıldığı görülmüştür.

4.6. Lojistik Regresyon Analizi Sonuçları

Lojistik regresyon analizinde literatürde de belirtildiği gibi varsayımlar yoktur. Bu sebeple analize doğrudan hiçbir varsayım araştırması yapılmadan başlanacaktır. Yapılacak analizde lojistik regresyon modeline değişkenlerin alınış yöntemlerine göre tüm değişkenlerin modele dâhil edildiği yöntem kullanılmıştır. Sonrasında modelde kullanılan değişkenlerin önem derecesinin tespit edilebilmesi ve daha az değişkenle sınıflandırma yapabilecek modelin elde edilebilmesi için ileri adımsal olabilirlik oranı (likelihood ratio) yaklaşımı kullanılmıştır. Sonuçlar ayrı ayrı gösterilerek yorumlanmıştır.

4.6.1.Tüm değişkenlerin modele dâhil edilmesi ile yapılan lojistik regresyon analizi

Kullanılan 155 ülkelik örneklemin 120'sinin lojistik regresyon analizinde kullanıldığı Çizelge 4.14'de görülmektedir. Lojistik regresyon analizinde de bağımlı değişken olarak çok gelişmiş ve orta düzeyde gelişmiş ülke sınıflandırması kullanılmıştır. Ayrıca bağımsız değişken olarak ta 16 bağımsız değişken analize alınmıştır.

Çizelge 4.14 İşleme Alınan Örneklem Sayısı

Ağırlıklandırılmamış Durumlar		N	Yüzde
Seçilen Durumlar	Analize Dâhil Edilen	120	77,4
	Kayıp Durumlar	35	22,6
	Toplam	155	100,0
	Seçilmeyen Durumlar	0	0
	Toplam	155	100,0

Başlangıç durumunda lojistik regresyon analizinde referans olarak hangi değerlerin alınacağı Çizelge 4.15'de belirtilmiştir. Araştırmada Çok Gelişmiş grup için 0 ve Orta Düzeyde gelişmiş için 1 kodlanmıştır. Bunun sebebi orta düzeyde gelişmiş ülkelere ait şans kriterinin maksimum şans kriterine eşit oluşudur. Bu sayede Orta düzeyde gelişmiş ülkelerin referans olarak seçilmesi ile sınıflandırma başarısı analiz başlangıcından itibaren yüksek tutulabilecektir.

Çizelge 4.15. Bağımlı Değişkenin Kodlanması

Orijinal Değer	İç Değer
Çok Gelişmiş	0
Orta Düzeyde Gelişmiş	1

Çizelge 4.16. İterasyon Geçmişi

İterasyon		-2 Log olabilirlik	Katsayılar
			Sabit
Adım 0	1	165,822	0,133
	2	165,822	0,134

4.6.1.1. Analiz sonucu elde edilen lojistik regresyon modeli

Çizelge 4.17. Model Katsayıları

İterasyon		-2 Log Olabilirlik	Katsayılar																
			Sabit	topnüf	Kentleşme Oranı	Kadın par	Saglık1	Saglık2	Saglık3	Yasam Bek	egitim3	iletisim1	iletisim2	iletisim3	GSYİH1	import	export	Elektrik tuk	Hapis
Adım 1	1	67,002	4,509	0,000	0,003	-0,002	-0,113	-0,025	0,000	-0,045	0,011	-0,004	-0,002	0,000	0,000	0,020	-0,020	0,000	0,000
	2	45,638	10,871	0,000	0,005	-0,014	-0,220	-0,089	0,001	-0,135	0,028	-0,006	-0,003	0,000	0,000	0,038	-0,043	0,000	0,000
	3	34,762	19,692	0,000	0,008	-0,026	-0,264	-0,138	0,000	-0,259	0,043	-0,007	-0,004	0,000	0,001	0,050	-0,061	0,000	0,000
	4	25,604	32,238	0,001	0,013	-0,031	-0,136	0,058	-0,001	-0,412	0,027	-0,008	-0,003	0,001	0,001	0,035	-0,048	0,000	0,000
	5	15,105	59,227	0,003	0,024	-0,027	0,040	0,651	-0,004	-0,678	-0,076	-0,007	0,000	-0,002	0,004	-0,009	0,007	-0,001	0,000
	6	8,407	98,192	-0,007	0,033	-0,037	0,119	1,328	-0,006	-1,032	-0,207	-0,010	0,002	-0,004	0,007	-0,062	0,063	-0,002	0,000
	7	4,111	160,414	-0,027	0,042	-0,053	0,046	2,150	-0,008	-1,592	-0,395	-0,019	0,003	-0,003	0,011	-0,140	0,133	-0,004	0,000
	8	1,624	243,438	-0,040	0,051	-0,068	-0,195	3,131	-0,009	-2,350	-0,635	-0,032	0,003	0,000	0,015	-0,227	0,208	-0,006	0,000
	9	0,609	331,648	-0,055	0,067	-0,083	-0,618	4,156	-0,011	-3,150	-0,895	-0,044	0,003	0,002	0,020	-0,306	0,275	-0,008	0,000
	10	0,226	419,238	-0,070	0,096	-0,104	-1,175	5,113	-0,013	-3,948	-1,155	-0,054	0,003	0,005	0,024	-0,365	0,327	-0,010	0,000
	11	0,083	504,386	-0,084	0,139	-0,135	-1,808	5,980	-0,015	-4,728	-1,410	-0,063	0,004	0,006	0,029	-0,401	0,363	-0,012	0,000
	12	0,031	587,833	-0,098	0,190	-0,172	-2,484	6,800	-0,018	-5,493	-1,664	-0,070	0,004	0,007	0,033	-0,425	0,390	-0,014	0,000
	13	0,011	670,633	-0,112	0,244	-0,212	-3,180	7,599	-0,021	-6,250	-1,918	-0,077	0,004	0,008	0,038	-0,443	0,414	-0,016	0,000
	14	0,004	753,228	-0,126	0,299	-0,253	-3,886	8,389	-0,024	-7,003	-2,174	-0,083	0,004	0,009	0,042	-0,458	0,435	-0,018	0,000
	24	0,000	1,586E3	-0,268	0,810	-0,634	-10,510	16,708	-0,053	-14,665	-4,683	-0,154	0,006	0,021	0,088	-0,683	0,705	-0,039	0,000

Çizelge 4.17'de görülebileceği gibi 24'üncü iterasyonda -2 log olabilirlik değerinin sıfır olması ve en uygun modelin bulunması sebebi ile iterasyonlar bu aşamada durdurulmuştur. Burada en önemli etkinin özel sağlık harcamaları, kamu sağlık harcamaları ve doğumda yaşam beklentisi tarafından yapıldığı dikkat çekicidir. Bazı değişkenlerin katsayıları ise ihmal edilecek derecede düşük hesaplanmıştır.

4.6.1.2. Modelin anlamlılığının test edilmesi

Çizelge 4.18. Model Katsayıları için Omnibus Testi

		Ki-Kare	Sd	Anl.
Adım 1	**Adım**	**165,822**	**16**	**5.727172177316E-27**
	Blok	**165,822**	**16**	**5.727172177316E-27**
	Model	**165,822**	**16**	**5.727172177316E-27**

Geleneksel Ki-Kare metodu kullanılarak modelin anlamlılığı bu aşamada test edilmektedir. Bağımlı değişkenin bağımsız değişkenler tarafından hep birlikte kullanılması ile test edilebilmesi durumu da burada test edilmektedir. Çizelge 4.18'deki değerler incelendiğinde analizin doğrudan enter yöntemi ile yapılması nedeniyle Adım, Blok ve Model Ki-kare değerlerinin aynı olduğu görülebilmektedir. Şayet adımsal (Stepwise) yöntem kullanılsaydı bu durum her adım için değişebilecekti. Yapılan analiz neticesinde önem derecelerinin 0,05'den küçük olması ($p<0,05$) nedeniyle modelin anlamlı olduğu söylenebilir.

Ayrıca Modelin uygunluğunun test edilmesinde Hosmer ve Lemeshow testi de kullanılmaktadır. Çizelge 4.19'da elde edilen değerler görülebilmektedir.

Çizelge 4.19. Hosmer ve Lemeshow Uyum İyiliği Testi Sonuçları

Adım	Ki-Kare	Sd	Anl.
1	**4.341199776598023E-9**	**5**	**1,000**

Hosmer ve Lemeshow testinde modelde tahmin edilen değerlerle gerçekte gözlenen değerler arasında fark yoktur sıfır hipotezi test edilmektedir.

H_0 : Teorik model verileri iyi temsil etmektedir.

H_A : Teorik model verileri iyi temsil etmemektedir.

Elde edilen önem derecesinin 1 olması sebebi ile (p>0,05) modelin tahmin ettiği verilerin istenen anlamlılık düzeyinde kabul edilebilir olduğu söylenebilir. Hosmer ve Lemeshow uyum iyilik testi model ile elde edilen tahminlerin gerçek gruplardan anlamlı bir fark oluşturmadığını göstermektedir. Bu, bağımsız değişkenin varyans açıklama yüzdesini göstermez ancak en azından açıklamanın anlamlı olduğunu ifade eder. Örneklem büyüklüğü arttıkça Hosmer ve Lemeshow uyum iyiliği testi daha hassas değerlerle daha küçük farkları bulabilir.

Çizelge 4.20. Modelin Özeti

Adım	-2 Log olabilirlik	Cox & Snell R Kare	Nagelkerke R Kare
1	7.007720948987392E-8	0,7488855877584233	0.9999999998041824000

Çizelge 4.20'de de görülebileceği gibi 25'inci iterasyonda en uygun model bulunmuştur. –2 Log olabilirlik istatistiği modelin ne kadar güçlü veya zayıf kararlar verebileceğini ifade etmektedir. Yüksek değerler alması durumunda modelin zayıf olduğu, çok küçük değerler aldığında ise modelin iyi olduğu belirtilmektedir. Çalışılan modelde –2 Log olabilirlik istatistiği yaklaşık 0 olduğundan modelin iyi olduğu söylenebilir.

Cox ve Snell R Kare istatistiği regresyondaki R kare değeri ile aynı yorumlanabilir. Yani bağımsız değişkenlerin bağımlı değişkeni %74,9 açıklayabildiği söylenebilir. Ancak Cox ve Snell R Kare istatistiği asla 1 değerini alamaz, bu durum da yorum yapmayı güçleştirebilir. Nagelkerke R Kare istatistiği ise Cox ve Snell R Kare istatistiğinin modifiye edilmiş halidir. Nagelkerke R Kare istatistiği genelde Cox ve Snell R Kare istatistiğinden yüksek bir değer almaktadır. Modelde Nagelkerke R Kare istatistiği yaklaşık 1'dir. Yani bağımsız değişkenlerin bağımlı değişkeni %100 açıklayabildiği söylenebilir.

4.6.1.3. Sınıflandırma sonuçları

Çizelge 4.21. Hosmer ve Lemeshow Kontenjans Tablosu

		Ülkelerin Gelişmişlik Sınıflandırması = Çok Gelişmiş		Ülkelerin Gelişmişlik Sınıflandırması = Orta Düzeyde Gelişmiş		
		Gözlenen	Beklenen	Gözlenen	Beklenen	Toplam
Adım 1	1	12	12,000	0	0,000	12
	2	12	12,000	0	0,000	12
	3	12	12,000	0	0,000	12
	4	12	12,000	0	0,000	12
	5	8	8,000	4	4,000	12
	6	0	0,000	8	8,000	8
	7	0	0,000	52	52,000	52

Çizelge 4.21'de Hosmer ve Lemeshow kontenjans tablosunda birinci adımda örneklemin 7 grup halinde oluşturularak yapılan tahminler görülebilmektedir.

Çizelge 4.22. Lojistik Regresyon Analizi Sınıflandırma Sonucu

		Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Geçerli	Çok Gelişmiş	56	36,1	46,7	46,7
	Orta Düzeyde Gelişmiş	64	41,3	53,3	100,0
	Toplam	120	77,4	100,0	
Kayıp	Sistem	35	22,6		
Toplam		155	100,0		

Çizelge 4.22'de tüm değişkenlerin modele dâhil edilmesi yapılan lojistik regresyon analizinde %100'lük bir sınıflandırma başarısı olduğu görülebilmektedir. Ayrıca uç değerlerin (outliers) analizi yapılmış ve uç değer bulunmadığı için örneklemden herhangi bir denek çıkarılmamıştır.

4.6.2. İleri adımsal olabilirlik yaklaşımı ile lojistik regresyon analizi

Lojistik regresyon analizi ile yüksek bir sınıflandırma oranı yakalanmasının yanı sıra modele etki eden önemli değişkenlerin neler olduğu ve bu değişkenler yardımıyla sınıflandırma başarısının ne olabileceği konusu yapılacak analizle ortaya konmaya çalışılacaktır. Sonuçta lojistik regresyon analizinde değişken sayısının azaltılarak

modelin yorumlanabilirliğinin kolaylaştırılması da ayrı bir amaç olarak değerlendirilmektedir.

İleriye adımsal olabilirlik yaklaşımında değişkenlerin modele alınmasında 0,15 ve çıkarılmasında 0,25 olasılığı kullanılmıştır. Bu analizde de yalnızca modele ilave edilen değişkenler ve sınıflandırma sonucu burada gösterilecektir.

Lojistik regresyon analizi sonucunda Çizelge 4.22'de bulunan değişkenler modelde kullanılmıştır. Denklemde; kentleşme oranı (kentoranı), Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2007) (Saglık1), Doğumda Yaşam Beklentisi (YasamBek), İlköğretime net kayıt oranı (Egitim3), 1000 kişiye düşen telefon hattı sayısı (2005) (İletisim1) ve Kişi başına Elektrik Tüketimi (elektriktuk) değişkenleri kullanılmıştır.

Çizelge 4.23. Modelde Kullanılan Değişkenler

		B	Standart Hata	Wald	sd	Anl.	Exp(B)	95,0% G.A. EXP(B)	
								Alt	Üst
Adım 9	Kentoranı	1,655	39,444	0,002	1	0,967	5,231	0,000	1,965E34
	Saglık1	-31,080	675,622	0,002	1	0,963	0,000	0,000	.
	YasamBek	-27,656	499,266	0,003	1	0,956	0,000	0,000	.
	egitim3	-9,243	171,574	0,003	1	0,957	0,000	0,000	1,072E142
	iletisim1	-0,342	9,688	0,001	1	0,972	0,710	0,000	1,252E8
	elektriktuk	-0,068	1,203	0,003	1	0,955	0,935	0,088	9,881
	Sabit	3076,479	54108,286	0,003	1	0,955	.		

Kullanılan değişkenler sonucunda aşağıdaki regresyon denklemi elde edilmiştir:

ln(Z)= 3076,479+1,655*kentoranı (Kentleşme Oranı)-**31,080***Saglık1 (GSYİH'nın yüzdesi olarak Kamu Sağlık Harcamaları(2005))-**27,656***YasamBek(Doğumda Yaşam Beklentisi(2002-2005))-**9,243***egitim3(İlköğretime Net Kayıt Oranı)-**0,342***İletişim1(1000 kişiye düşen telefon hattı sayısı (2005)) - **0,068***elektriktuk(Kişi Başına Elektrik Tüketimi (Kw-H olarak)(2004))

Denkleme 9'uncu adımda değişkenler dâhil edilmiş ve bu değişkenler kullanılarak sınıflandırma sonucu bulunmuştur. Değişkenler incelendiğinde kentleşme oranının pozitif bir etki yarattığı, diğer sağlık, eğitim, iletişim ve enerji değişkenlerinin negatif etki yarattığı gözlenmektedir. Başlangıçta sayıca çok olan grubun şans kriteri yüksek olduğu için 1 olarak belirlendiği ifade edilmişti. Yani çok gelişmiş ülke

grubunda model 0 ve 0'a yakın sonuçları dikkate alacak, orta düzeyde gelişmiş grupta ise 1 ve 1'e yakın değerleri dikkate alacaktır. Bu açıklama ile denklem katsayılarının özellikle negatif etki yaratanların çok gelişmiş ülke grubuna sınıflandırmada etkili olduğu söylenebilir. En yüksek etkinin iletişim1 ve kişi başına elektrik tüketimi değişkenleri tarafından olduğu görülebilmektedir.

Yapılan incelemede kurulan modelin uyum iyiliği ve değişkenlerin anlamlılığı test edilmiş sonuçları olumlu çıkmıştır (9'uncu Adımda -2 Log olabilirlik=0, Cox-Snell R Kare=0,749; Nagelkerke R Kare=1; Hosmer Lemeshow Testi Ki-Kare=0; Sd=3; $p<0,05$). Sınıflandırma sonucunda ise Çizelge 4.24'de görülebileceği gibi %100'lük bir başarı elde edilmiştir.

Çizelge 4.24. İleri Adımsal Olabilirlik Yaklaşımı Sınıflandırma Sonucu

			Tahmin Edilen		
			Ülkelerin Gelişmişlik Sınıflandırması		
	Gözlenen		Çok Gelişmiş	Orta Düzeyde Gelişmiş	Yüzdesi Doğru
Adım 9	Ülkelerin Gelişmişlik Sınıflandırması	Çok Gelişmiş	56	0	100,0
		Orta Düzeyde Gelişmiş	0	64	100,0
		Toplam Yüzdesi			100,0

4.7.Yapay Sinir Ağları Analizi ve Sonuçları

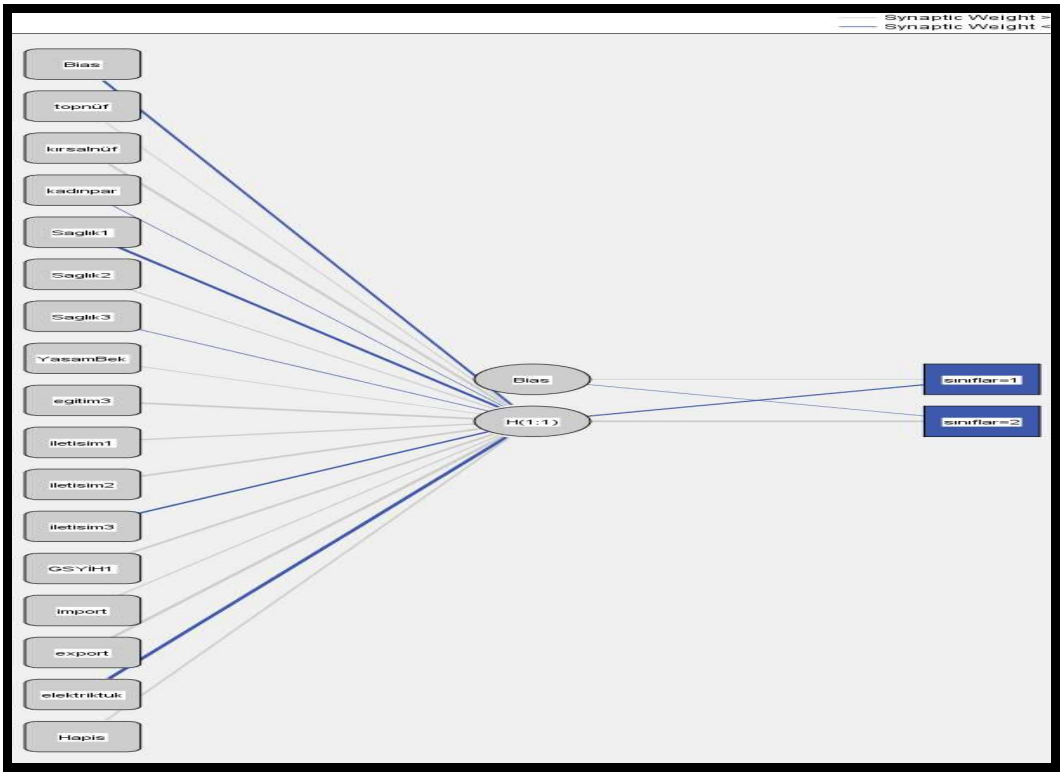
Yapay sinir ağları uygulamasında öncelikle verilerin herhangi bir ön işleme tabi tutulmadan modele dâhil edilmesi sağlanmış ve sınıflandırma sonuçları alınmış, bilahare verilerin standartlaştırılması sonucu model oluşturularak sınıflandırma sonuçları açısından karşılaştırılmıştır. Yapay sinir ağı uygulamasında başlangıçta çok katmanlı perseptron modeli ve radyal tabanlı ağ kullanılmıştır. Radyal tabanlı ağdan elde edilen sonuçların tatminkâr olmaması sebebi ile çalışmaya çok katmanlı perseptron modelleri ile devam edilmiştir. Çizelge 4.25'de işleme alınan örneklem sayısı belirtilmiştir. Buna göre ağa sınıflandırılacak ülkelere ilişkin verilerin eğitilmesi maksadıyla 71 ülkeden oluşan örneklem, modelin testi için 36 ülkeden oluşan örneklem ve modelin geçerlemesinin yapılması için ise 13 ülkeden oluşan örneklem tahsis edilmiştir.

Çizelge 4.25. İşleme Alınan Örneklem Sayısı

		N	Yüzde
Örneklem	Eğitim	71	59,2%
	Test	36	30,0%
	Alıkoyma	13	10,8%
Geçerli		120	100,0%
Çıkarılan		35	
Toplam		155	

16 değişken giriş katmanına dâhil edilmiş, dâhil edilen değişkenler ham veri olarak kullanılmış, bir adet gizli katman kullanılmış, aktivasyon fonksiyonu olarak hiperbolik tanjant fonksiyonu kullanılmıştır. Çıktı aktivasyon fonksiyonu olarak Softmax kullanılmıştır. Bu aşamada verilerin herhangi bir işlemden geçirilmeden analize dâhil edilmek suretiyle de yapay sinir ağları ile oluşturulacak modellerin veri yapısına olan hassasiyeti ölçülmek istenmektedir. Birçok farklı girdi ve çıktı aktivasyon fonksiyonu denenmiş ancak en iyi sonucu veren fonksiyonlar kullanılarak analiz çalışmasına devam edilmesine karar verilmiştir.

Şekil 4.5'te ağ yapısı ile birlikte 16 değişken ve ikili bağımlı değişken şematik olarak gösterilmektedir. Modelde bir adet gizli katman kullanılmıştır.



Şekil 4.5. Ham Verilerle Kurulan Çok Katmanlı Ağ Yapısı

Çizelge 4.26'da Yapay sinir ağları ile yukarıda belirtilen ağ mimarisine karşılık analiz yapılmış ve analiz sonucunda eğitim örnekleminde %98,6'lık, test örnekleminde %97,2'lik, alıkoyma örnekleminde de %84,6'lık bir sınıflandırma başarısı elde edilmiştir. Ancak bu sonuçların ham veri ile ölçeklendirme yapılmadan verilerin girdi katmanına dahil edilmesi sonucu oluştuğu düşüncesiyle değişkenler normalleştirilerek analiz tekrarlanmıştır.

Çizelge 4.26. Sınıflandırma Tablosu

Örneklem	Gözlenen	Tahmin Edilen		
		Çok Gelişmiş	Orta Düzeyde Gelişmiş	Yüzde Doğru
Eğitim	Çok Gelişmiş	30	1	96,8%
	Orta Düzeyde Gelişmiş	0	40	100,0%
	Toplam Yüzde	42,3%	57,7%	98,6%
Test	Çok Gelişmiş	17	1	94,4%
	Orta Düzeyde Gelişmiş	0	18	100,0%
	Toplam Yüzde	47,2%	52,8%	97,2%
Alıkoyma	Çok Gelişmiş	5	2	71,4%
	Orta Düzeyde Gelişmiş	0	6	100,0%
	Toplam Yüzde	38,5%	61,5%	84,6%

Normalleştirme sonucunda Şekil 4.6'da bulunan ağ mimarisi geliştirilmiş ve Çizelge 4.27'de bulunan sınıflandırma tablosu elde edilmiştir.

Şekil 4.6. Normalize Edilen Verilerle Kurulan Çok Katmanlı Ağ Yapısı

Şekil 4.6'da bulunan ağ yapısında da yine 1 girdi katmanı 1 gizli katman ve 1 çıktı katmanı kullanılmıştır.

Çizelge 4.27. Sınıflandırma Tablosu

Örneklem	Gözlenen	Tahmin Edilen		
		Çok Gelişmiş	Orta Düzeyde Gelişmiş	Yüzde Doğru
Eğitim	Çok Gelişmiş	26	1	96,3%
	Orta Düzeyde Gelişmiş	3	35	92,1%
	Toplam Yüzde	44,6%	55,4%	93,8%
Test	Çok Gelişmiş	20	2	90,9%
	Orta Düzeyde Gelişmiş	2	14	87,5%
	Toplam Yüzde	57,9%	42,1%	89,5%
Alıkoyma	Çok Gelişmiş	7	0	100,0%
	Orta Düzeyde Gelişmiş	0	10	100,0%
	Toplam Yüzde	41,2%	58,8%	100,0%

Çizelge 4.27'de yapay sinir ağları ile Şekil 4.6'da belirtilen ağ mimarisine karşılık analiz yapılmış ve analiz sonucunda eğitim aşamasında %93,8'lik, test aşamasında %89,5'lik, geçerleme aşamasında da %100'lük bir sınıflandırma başarısı elde edilmiştir.

Yapılan başarılı sınıflandırma sonucunda değişkenlerin yüzde olarak önemlilikleri ise Çizelge 4.28'de gösterilmektedir.

Çizelge 4.28. Bağımsız Değişken Önem Yüzdesi Çizelgesi

	Önem	Normalleştirilmiş Önem
Toplam Nüfus(2005)	0,010	3,6%
Kentleşme Oranı(2005)	0,003	1,1%
Kadın Parlamenter Oranı (Toplamın Yüzdesi)	0,003	1,1%
Sağlık Harcamaları Kamu (GSYİH'nın yüzdesi)(2004)	4,663E-9	1,7E-6%
Sağlık Harcamaları Özel (GSYİH'nın yüzdesi)(2004)	7,467E-7	2,7E-4%
Sağlık Harcamaları Kişi Başına (Satın Alma Gücü Paritesine göre US$) (2004)	0,069	25,4%
Doğumda Yaşam Beklentisi(2002-2005)	0,004	1,6%
İlköğretime net kayıt oranı	2,131E-5	,0%
1000 kişiye düşen telefon hattı sayısı (2005)	0,091	33,5%
1000 kişiye düşen cep telefonu abonesi sayısı (2005)	0,086	31,7%
1000 kişiye düşen internet kullanıcısı sayısı (2005)	0,039	14,3%
GSYİH (Milyar Dolar) (2005)	0,145	53,3%
İthal Edilen Mallar (GSYİH %'si olarak) (2005)	0,016	6,0%
İhraç Edilen Mallar (GSYİH %'si olarak) (2005)	0,021	7,6%
Elektrik Tüketimi (Kw-H olarak)(2004)	0,272	100,0%
Hapiste Bulunan Kişi Sayısı (2007)	0,242	89,0%

Çizelge 4.28'de de görülebileceği gibi kurulan çok katmanlı perseptron modelinde kişi başına elektrik tüketimi en yüksek öneme sahipken, hapiste bulunan kişi sayısı, GSYİH, 1000 kişiye düşen telefon hattı sayısı, 1000 kişiye düşen cep telefonu abonesi sayısı gibi değişkenler önem sırasını takip etmektedirler. Kalkınmışlık için önemli olduğu model tarafından belirtilen değişkenler gerçekte de bir ülkenin refah düzeyinin belirlenmesinde önemli faktör olarak değerlendirilebilir. İlginç olan önem derecesi en yüksek olan değişken olarak elektrik tüketiminin ve bir sonraki öneme sahip olan değişkenin ise hapiste bulunan kişi sayısının belirlenmiş olmasıdır. Belirtilen iki değişkeninde sınıflandırmada ayırt edicilik açısından model içerisinde yüksek seviyede bir öneme haiz olduğu bu analiz ile görülebilmektedir.

Yapay sinir ağları modellerinin performanslarının analizinde ayrıca ROC analizleri yapıldığı da bilinmektedir. Hatta Türker v.d. (2005) yaptıkları analizde kurulan 10 modelde sınıflandırma başarısı en iyi olan modelin öğrenme algoritmasının ROC analizi sonucunda istenen düzeyde olmadığı bu yüzden sınıflandırma başarısı ve ROC analizi en iyi olan 2 modelin doğru teşhis için kullanılabileceği ifade edilmiştir. ROC analizlerinde eğri altında kalan alan şayet 0,5'e eşitse tahmin olasılığı eşit olduğundan ayrımın iyi yapılamadığı, 0,7 ila 0,8 ise ayırma yeteneğini istatistiksel olarak kabul edilebilir, 0,8 ila 0,9 arasında ise ayırma yeteneği istatistiksel olarak mükemmel, 0,9'dan büyükse ayırma yeteneği istatistiksel olarak olağanüstü olarak değerlendirmektedir (Bayru 2007: 174).

Yapılan analizde yapay sinir ağı modeline ilişkin ROC eğrisi altında kalan alana ilişkin değerler Çizelge 4.29'da görülmektedir.

Çizelge 4.29. ROC Analizi Sonucu Eğri Altında Kalan Alan

Eğri Altında Kalan Alan		
		Alan
Ülkelerin Gelişmişlik Sınıflandırması	**Çok Gelişmiş**	**0,989**
	Orta Düzeyde Gelişmiş	**0,989**

Çizelge 4.29'da da görülebileceği gibi 0,989 ile istatistiksel olarak olağanüstü bir ayırma yeteneği model tarafından sergilenmiştir. Bu analizle modelin performansı da test edilmiştir.

SONUÇ

Sınıflandırma sosyal bilimlerde yapılan araştırmalarda birçok metot kullanılarak yapılmakta ve yapılan bu sınıflandırma sonuçlarına göre istatistiksel olarak çıkarımlar ortaya konmaktadır. Bu çalışmada sınıflandırma metotlarından Diskriminant Analizi, Lojistik Regresyon Analizi ve Yapay Sinir ağları modelleri kullanılarak örnek bir veri seti üzerinde uygulama yapılmıştır.

Kullanılan veri seti, ülkelerin Beşeri Kalkınma Endeksine göre Birleşmiş Milletler tarafından yapılan çok gelişmiş ve orta düzeyde gelişmiş ülke sıralamasında yararlandığı verilerdir. Bu çalışmada endeks değerleri yerine endeksleri oluşturan ham veriler kullanılmış ve ülkeler yeniden sınıflandırılmıştır.

Diskriminant analizinde normallik varsayımı, kovaryans matrislerinin eşitliği varsayımı ve çoklu bağlantı varsayımı sınanmıştır. Normallik varsayımının sınanmasında bazı değişkenlerin tek değişkenli normal dağılım göstermediği için normallik dönüşümleri yapılmış ve bilahare yapılan testlerde çok değişkenli normal dağılım gösterdiği gösterilmiştir. Kovaryans matrisleri eşitliği varsayımı sağlanamadığından doğrusal diskriminant analizi yerine kuadratik diskriminant analizi kullanılmıştır. Çoklu bağlantı probleminin olmadığı yapılan analizler neticesinde gösterilmiştir. Diskriminant analizi sonucunda bağımsız değişkenlerin bağımlı değişkeni %70,39 açıklayabildiği görülmüştür. Kullanılan bağımsız değişkenlerin önem değerlendirmesinde iletişim verilerinin, doğumda yaşam beklentisinin, GSYİH'nın yüzdesi olarak ihraç edilen malların, kamu sağlık harcamalarının çok gelişmiş ülke olabilmede pozitif etki yarattığı tespit edilmiştir. Ülkelerin gelişmişliklerinin belirlenmesinde bu durumun gerçekçi olduğu düşünülmektedir. Zira gelişmiş ve üretim yapan ülkelerde ihracatın yüksek olması, iletişimin daha iyi tesis edilmiş olması, sağlık yatırımlarının ve buna bağlı olarak doğumda yaşam beklentisinin üst seviyede görülmesi olağan bir durumdur. Kırsal nüfus, ithal edilen mallar ve ilköğretime net kayıt oranı değişkenlerinin ise modele negatif etkide bulundukları da tespit edilmiştir. Bu durumun da rasyonel olduğu değerlendirilmektedir. Çok gelişmiş ülke konumunda bir ülkenin sınıflandırılmasında diskriminant analizi sonuçları dikkate alındığında ihracat, iletişim ve sağlık yatırımlarının artırılmasının önemli olduğu açıktır. Yani

sağlıklı bireylerin koordineli bir şekilde el ele vererek üretime dönük çalışması bir ülkeyi çok gelişmiş ülke kategorisine taşıyabilecektir.

Diskriminant analizi ile yapılan sınıflandırmada %92,5'lik bir sınıflandırma başarısı elde edilmiştir. Hatalı sınıflandırılan ülkeler incelendiğinde bu ülkelerin Birleşmiş Milletler tarafından yapılan sıralamada çok gelişmiş ülkeler ve orta düzeyde gelişmiş ülkeler sınırına yakın ülkeler arasında olduğu görülmektedir. Bu durum hazırlanan modelin Birleşmiş Milletler Kalkınma Programı tarafından yapılan sıralamaya çok aykırı sonuçlar ortaya koymadığına da işaret etmektedir. Hatalı sınıflandırılan ülkelerden birisi de Türkiye'dir. Kullanılan ham verilere göre kurulan model Birleşmiş Milletler tarafından orta düzeyde gelişmiş ülkeler kategorisinde sınıflandırılan Türkiye'yi çok gelişmiş kategoride sınıflandırmaktadır. Bu durumunda önemli olduğu düşünülmektedir.

Diskriminant analizi ayrıca adımsal olarak yapılmış ve ayırt edici değişkenlerin neler olduğu konusu detaylı olarak incelenmiştir. Bu analiz sonucunda 4 değişken ile model oluşturulmuştur. Kamu sağlık harcamaları ve özel sağlık harcamalarının modelde negatif etki yarattığı görülmüş, satın alma gücü paritesine göre kişi başına sağlık harcamaları ve 1000 kişiye düşen cep telefonu abone sayısı da pozitif etkide bulunmuştur. Burada dikkati çeken en önemli konu yüksek etkisiyle satın alma gücü paritesine göre kişi başına düşen sağlık harcamalarıdır.

Lojistik regresyon analizinde ise başlangıçta tüm bağımsız değişkenlerin modele dâhil edilmesi ile diskriminant analizi ile yapılacak karşılaştırmada sonucun daha tutarlı yorumlanabileceği düşünülmüştür. Lojistik regresyon analizi ile amaçlanan daha az değişken kullanılarak sınıflandırmanın yapılabilmesi olduğundan ileri adımsal olabilirlik oran yaklaşımı kullanılarak yeni regresyon modelleri oluşturulmuştur. Kurulan bu adımsal modeller yardımıyla bağımsız değişken sayısının indirgenmesi ve modelin kolayca yorumlanması sağlanmıştır.

Tüm verilerin kullanıldığı regresyon modelinde özel sağlık harcamaları pozitif etki gösterirken, kamu sağlık harcamaları, yaşam beklentisi ve ilköğretime net kayıt oranı değişkenleri negatif etki göstermişlerdir. Başlangıçta nispi şans kriterinin yüksek tutulması için orta düzeyde gelişmiş grup "1" ile çok gelişmiş grup ise "0" ile kodlanmıştır. Yani Yaşam beklentisi, ilköğretime net kayıt oranı ve kamu sağlık

harcamaları değeri 0'a yaklaştırarak ülkenin çok gelişmiş sınıfına dâhil edilmesini sağlamaktadır.

İleri adımsal yaklaşımla oluşturulan lojistik regresyon modelinde 7 değişken seçilmiştir. Bu değişkenler genel olarak incelendiğinde modelin sağlık, eğitim, enerji ve iletişim değişkenlerinden oluştuğu gözlenmektedir. Ayrıca kentleşme oranında önemli ancak ters yönlü bir etkisi olduğu da gözlenmiştir. Yedi değişken ile de yapılan sınıflandırma neticesinde %100'lük bir sınıflandırma başarısı elde edilmiştir.

Yapay sinir ağları ile sınıflandırmada sıklıkla kullanılan çok katmanlı multi perseptron modeli kullanılmıştır. Kullanılan modelde girdi değişkenleri ham veri olarak ve normalize edilmiş olarak ayrı ayrı denenmiştir. Yapılan bu denemeler sonucunda ham verilerle yapılan analizde alıkoyma örneklem setinde %84,6'lık bir sınıflandırma başarısı elde edilmiş, Normalize edilmiş verilerin kullanılması durumunda ise %100'lük bir sınıflandırma başarısı elde edilmiştir. Bağımsız değişkenlerin önem dereceleri dikkate alındığında ise elektrik tüketimi, hapiste bulunan kişi sayısı, GSYİH ve iletişim değişkenlerinin yüzde olarak önemli bir öneme sahip olduğu görülmüştür. Çok gelişmiş ülkeler ile orta düzeyde gelişmiş ülkeler arasında belirtilen değişkenler arasında önemli bir farkın bulunması da beklenen bir durumdur. Bu analizde ilginç olabilecek bir konu ise ilköğretime net kayıt oranının, kamu sağlık harcamalarının ve özel sağlık harcamalarının yüzde olarak önem düzeyinin çok düşük bir değer almasıdır. Bu durumun yapay sinir ağları ile oluşturulan modelde diğer değişkenlerin daha baskın bir rol oynayarak bahse konu değişkenleri etkisizleştirdiği olarak yorumlanmıştır.

Sonuçta yapılan üç analizin de başarılı olduğu görülmektedir. Diskriminant analizinin varsayımlarının diğer yöntemlere göre daha fazla olmasının sınıflandırma başarısını olumsuz etkilediği düşünülmektedir. Lojistik regresyon analizi ve yapay sinir ağları ile kurulan modellerin varsayımlara bağlı kalmamasının %100'lük bir sınıflandırma başarısı yakalanmasına sebep olduğu değerlendirilmektedir. Analiz yöntemlerinin kullandıkları matematiksel modeller ve sınıflandırma problemine yaklaşım metotları değerlendirildiğinde katsayıların farklılık göstermesi beklenen bir durum olarak görülmektedir. Bu sebeple karşılaştırmada sadece sınıflandırma başarılarının dikkate alınması sonuçların karşılaştırılması açısından önemli görülmüştür.

Bu çalışma ile çok gelişmiş ülkeler ve orta düzeyde gelişmiş ülkelerin yeniden sınıflandırılması sonucunda kullanılan tüm modellerin sınıflandırma başarısı çok

yüksektir. Bu kapsamda sınıflandırmada makro veya mikro düzeylerde bu yöntemlerin tezde işaret edilen değişkenler kullanılarak yapılabileceği de gösterilmiştir. Bundan sonraki çalışmalarda bu tezde kullanılan veri seti ile değişkenler arasındaki ilişki kanonik korelasyon analizi ile incelenebilir ve bilahare değişkenler arasındaki ilişkiler daha da detaylandırılarak az gelişmiş ve orta düzeyde gelişmiş ülkelere beşeri kalkınma düzeylerini artırmada stratejik bir yol haritası teklif edilebilir.

KAYNAKLAR

Aasheım, C. ve Koehler, G.J., Scanning World Wide Web With the Vector Space Model, Decison Support Systems. 42, 690-699, 2006.

Ainscough, T.L. ve Aronson, J.E., An Empirical Investigation and Comparison of Neural Networks and Regression for Scanner Data Analysis, Journal of Retailing and Consumer Services, 6(4), 205-217, 1999.

Akgül, A., Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri, Yükseköğretim Kurulu Matbaası, Ankara, 1997.

Akgül, A. ve Çevik, O., İstatistiksel Analiz Teknikleri: SPSS'te İşletme Yönetimi Uygulamaları, Emek Ofset, Ankara, 1–85, 2003.

Akın, M., Türkiye'deki Enflasyonun Tahmini için Yapay Sinir Ağı ile Bir Uygulama, Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 29, 1997

Albayrak, A.S. Uygulamalı Çok Değişkenli İstatistik Teknikleri, Asil Yayın Dağıtım A.Ş. Ankara, 2006.

Allen, M.W., Ng S.H., Leiser D., Adult Economic Model and Values Survey:Cross-National Differences in Economic Beliefs, Journal of Economic Psychology, 26, 159-185, 2005.

Balcaen, S. ve Ooghe, H., 35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems, The British Accounting Review. 38, 63-93, 2006.

Baş, N., Yapay Sinir Ağları Yaklaşımı ve Bir Uygulama. Yüksek Lisans Tezi, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, 2006.

Başarır, G., Çok Değişkenli Verilerde Ayrımsama Sorunu ve Lojistik Regresyon Analizi, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 1990.

Bayru, P., Elektronik Basında Tüketici Tercihleri Analizi: Yapay Sinir Ağları ile Lojit Modelin Performans Değerlendirilmesi, Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 2007.

Berg, D., Bankruptcy Prediction by Generalized Additive Models, Applied Stochastic Models in Business and Industry, 23, 129-143, 2004.

Bhanoji, R., Human Development Report:1990 Review Assesment, World Development, 19, 10, 1991.

Bonney, G.E., Logistic Regression for Dependent Binary Observations, Biometrics, 43, 951-973, 1987.

Bosse, D.A., Bundling Governance Mechanisms to Efficiently Organize Small Firm Loans, Journal of Business Venturing, 24, 183-195, 2008.

Bressan, M. ve Vitria, J., Nonparametric Discriminant Analysis and Nearest Neighbour Classification, The Journal of The Pattern Recognition Society, 24, 2743-2749, 2003.

Cangül, A., Diskriminant Analizi ve Bir Uygulama Denemesi, Uludağ Üniv. Sosyal Bilimler Enstitüsü Ekonometri Ana Bilim Dalı, 2006.

Canyüker, E. ve Aşan, Z., Parametrik Olmayan İstatistik Teknikler, Anadolu Üniversitesi Yayınları. Eskişehir, 2005.

Chen, K., Yen, D.C., Hung, S., Huang, A.H., An Exploratory Study of the Selection of Communication Media: The Relationship Between Flow and Communication Outcomes, Decision Support Systems, 45, 822-832, 2008.

Cheng, B. ve Tıtterıngton, D.M., Neural Networks: A Review From a Statistical Perspective, Statistical Science, 9(1), 2-30, 1994.

Conover, W.J., Practical Nonparametric Statistics, Wiley Publication, 1999.

Cornfıeld, J., Joint Dependence Of The Risk Of Coronary Heart Disease On Serum Cholesterol And Systolic Blood Pressure: A Discriminate Function Analysis., Federation Proceedings, 21, 58-61, 1962.

Çamdeviren, H., Lojistik Regresyon ve Diskriminant Analizi, Doktora Tezi, Ankara Üniversitesi, Ankara, 89-91, 2000.

Çilan, Ç.A., Bolat, B.A., Coşkun, E., Analyzing Digital Divide Within and Between Member and Candidate Countries of European Union, Government Information Quarterly, 26, 98-105, 2009.

Düzgüneş, O., Kesici, T., Kavuncu, O., Gürbüz, F., Araştırma ve Deneme Metotları, Ankara Üniversitesi Ziraat Fakültesi Yayınları, Ankara, 1987.

Efe, Ö. ve Kaynak, O., Yapay Sinir Ağları ve Uygulamaları, Boğaziçi Üniversitesi Yayınları, İstanbul, 2004.

Elmas, Ç., Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama), Seçkin Yayıncılık. Ankara, 2003.

Emel, A.B., Oral, M., Reisman, A., Yolalan, R., A Credit Scoring Approach for the Commercial Banking Sector, Socio-Economic Planning Sciences, 37, 103-123, 2003.

Erçetin, Y., Diskriminant Analizi ve Bankalar Üzerine Bir Uygulama, Türkiye Kalkınma Bankası A. Ş. APM/28 (KİG-26), 1-2, 1993.

Fukunaga, K. ve Mantock, J. Nonparametric Discriminant Analysis, IEEE Trans. PAMI 5, 671-678, 1983.

Gan, C., Limsonbuchai, V., Clemes, M., Weng, A., Consumer Choice Prediction: Artificial Neural Network Versus Logistic Models, Journal of Social Sciences, 1(4), 211-219, 2005.

Gao, H. ve Davis, J.W., Why Direct LDA is not Equivalent to LDA?, The Journal of The Pattern Recognition Society, 39, 1002-1006, 2005

Gaspar, D., Kalkınma Ahlakı Yeni Bir Alan mı? Piyasa Güçleri ve Küresel Kalkınma. İngilizceden çeviren: İdil Eser. İstanbul: Yapı Kredi Yayınları No:2, 1995.

Gujarati, D.N., Temel Ekonometri. Çev.Ümit ŞENESEN, Gülay G.ŞENESEN. İstanbul, 2001.

Hair, J., Anderson, R. E., Tahtam, R.L., Black, W.C., Multivariate Data Analysis, Prentice Hall, New Jersey, Fifth Edition, 1995.

Han, E. VE Kaya, A. Kalkınma Ekonomisi, Teori ve Politika, İkinci Basım, Eskişehir: Birlik Ofset, 1997.

Hastie, T., Buja, A., Tibshirani, R., Penalized Discriminant Analysis, The Annals of Statistics, 23:1, 73-102, 1995.

Hertz, J., Krogh, A., Palmer, R.G., Introduction The Theory of Neural Computation. Addison-Wesley Publishing, 1991.

Holden, M.T., O'toole, T., A Quantitative Exploration of Communication's Role in Determining the Governance of Manufacturer-Retailer Relationships, Industrial Marketing Management, 33, 539-548, 2004.

Hosmer, D.W. ve Lemeshow, S., Applied Logistic Regression, John Wiley &Sons, 1998.

Hosmer, D.W., Taber, S., Lemeshow, S., The Importance of Assessing the Fit of Logistic Regression Models: A Case Study, American Journal of Public Health, 81, 1630-1639, 1991.

Hwang, H., Ku, C., Yen, D., Cheng, C., Critical Factors Influencing the Adoption of Data Warehouse Technology: A Study of Banking Industry in Taiwan, Decison Support Systems, 37, 1-21, 2004.

Kaşko, Y., Çoklu Bağlantı Durumunda İkili (Binary) Lojistik Regresyon Modelinde Gerçekleşen I.Tip Hata ve Testin Gücü, Yüksek Lisans Tezi Ankara Üniversitesi Zootekni Ana Bilim Dalı, 2007.

Klecka, W., Discriminant Analysis, Sage Publications, London, 1980.

Lachenbruch, P.A., Discriminant Analysis, Hafner Pres, London, 1975.

Landajo, M. A. ve Lorcha, P., Robust Neural Modeling for the Cross-Sectional Analysis of Accounting Information, European Journal of Operational Research, 177(2), 1232-1252, 2007.

Lattin, J., Carroll, J.D., Gren, P.E., Analyzing Multivariate Data, Brooks/Cole Thomson, Canada, 2003.

Lee, K., Booth, D., Alam, P., Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Firms, Expert Systems With Applications, 29(1), 1-16, 2005.

Lee, C.T., Logistic Models for Cross-over Designs, Biometrika, 71, 216-217, 1984.

Leshno, M. ve Spector, Y., Neural Network Prediction Analysis: The Bankruptcy Case, Neurocomputing, 10, 125-147, 1996.

Liang, Z. ve Shi, P., Kernel Discriminant Analysis and Its Theoretical Foundation, The Journal of The Pattern Recognition Society. 38, 445-447, 2004.

Liang, Y., Gong, W., Pan, Y., Li, W., Generalizing Relevance Weighted LDA, The Journal of The Pattern Recognition Society, 38, 2217-2219, 2005.

Li, S. ve Lin, B., Accessing Information Sharing and Information Quality in Supply Chain Management, Decision Support Systems, 42, 1641-1656, 2006.

Limsombunchai, V., Gan, C., Lee, M., An Analysis of Credit Scoring for Agricultural Loans in Thailand, American Journal of Applied Sciences, 2(8), 1198-1205, 2005.

Lu, J. Plataniotis, K.N., Venetsanapoulos, A.N., Wang, J., An Efficient Kernel Discriminant Analysis Method, The Journal of The Pattern Recognition Society, 38, 1788-1790, 2005.

Ma, B., Qu, H., Wong, H., Kernel Clustering-Based Discriminant Analysis, The Journal of Pattern Recognition Society, 40, 557-562, 2007.

Malhotra, R. ve Malhotra, D.K., Evaluating Consumer Loans Using Neural Networks, The International Journal of Management Sciences, 31, 83-96, 2003.

Mardia, K.V., Kent, J.T., Bibby, M.J., Multivariate Analysis, Academic Pres Limited, USA, 300-325, 1979.

Mclahlan, G.J., Discriminant Analysis and Statistical Pattern Recognition, John Wiley and Sons Publication, New Jersey, USA, 2004.

Odom, M. ve Sharda, R., A Neural Network Model for Bankruptcy Prediction, Proceedings of the International Joint Conference on Neural Networks, IEEE Press, CA, 2, 163-168, 1990.

Öztemel, E., Yapay Sinir Ağları, Papatya Yayıncılık, İstanbul, 2003.

Öztürk, A. 2006. Esnek Ayırma Analizi ve Bir Uygulama. Eskişehir Osmangazi Üniv. Biyoistatistik Bilim Dalı.

Park, C. ve Park, H., A Comparison of Generalized Linear Discriminant Analysis Algorithms, Pattern Recognition, 41, 1083-1097, 2007.

Pasiouras, F. ve Tanna, S., The Prediction of Bank Acquisition Targets with Discriminant and Logit Analysis:Methodological Issues and Empirical Advances, Research in International Business and Finance, Doi:10.1016/j.ribaf.2009.01.004, 2009.

Pollalis, Y.A., Patterns of Co-alignment in Information-Intensive Organizations: Business Performance Through Integration Strategies, International Journal of Information Management, 23, 469-492, 2003.

Pompe, P.P.M. ve Bilderbeek, J., The Prediction of Bankruptcy of Small-and-Medium Sized Industrial Firms, Journal of Business Venturing, 20, 847-868, 2005.

Rachel, S.D., Towards a Better Understanding of Partnership Attributes: An Exploratory Analysis of Relationship Type Classification, Industrial Marketing Management, 37, 228-244, 2008.

Rasson, J.P. ve Bertholet, V., Application of the Poisson Process Model for The Early Detection of Enterprise Bankruptcy, Applied Stochastic Models in Business and Industry, 15, 443-449, 1999.

Reynes, C., Sabatier, R., Molinari, N., Choice of B-Splines With Free Parameters in the Flexible Discriminant Analysis Context, Computational Statistics&Data Analysis, 51, 1765-1778, 2006.

Roberts, G., Rao, J.N.K., Kumar, S., Logistic Regression Analysis of Sample Survey Data, Biometrika, 74(1), 1-12, 1987.

Schmid, U., Rosch, P., Krause, M., Harz, M., Popp, J., Baumann, K., Gaussian Mixture Analysis for the Single-Cell Differentiation of Bacteria Using Micro-Raman Spectroscopy, Chemometrics and Intelligent Laboratory Systems, 96, 159-171, 2009.

Sharma, S., Applied Multivariate Techniques, John Wiley and Sons Inc, Canada, 1996.

Sharma, S., Rai, A., An Assessment of the Relationship Between ISD Leadership Characteristics and IS Innovation Adoption in Organizations, Information Management, 40, 391-401, 2003.

Shin, K., SPSS Guide for DOS Version 5 and Windows 6.1.2, 2nd Ed. Chicago, 1996.

Sheth, J.N., The Multivariate Revolution in Marketing Research, Journal of Marketing. 35, 13-19, 1971.

SPSS Regression Models 16.0., www.hanken.fi/student/media/3618/ spssregressionmodels160. pdf, Erişim Tarihi: 04.09.2009.

Srivastava, S., Gupta, M., Frigyik, B., Bayesian Quadratic Discriminant Analysis, Journal of Machine Learning Research, 8, 1277-1305, 2007.

Sueyoshi, T., A Methodological Comparison Between Standard and Two Stage Mixed Integer Approaches for Discriminant Analysis, Asia-Pacific Journal of Operations Research, 4, 513-528, 2004.

Sueyoshi, T. ve Hwang, S., A Use of Non-Parametric Tests for DEA-Discriminant Analysis: A Methodological Comparison, Asia-Pacific Journal of Operational Research, 2004.

Tabachnick, B.G. ve Fidell, L.S., Using Multivariate Statistics Fourth Edition, Pearson Education Company, USA, 2001.

Tang, E.K., Suganthan, P.N., Yao, X., Qin, A.K., Linear Dimensionality Reduction Using Relevance Weighted LDA, The Journal of The Pattern Recognition Society, Cilt:38, 485-493, 2004.

Tang, H., Fang, T., Shi, P., Laplacian Discriminant Analysis, The Journal of The Pattern Recognition Society, 39, 136-139, 2005.

Tao, Q., Wu, G., Wang, J., The Theoretical Analysis of FDA and Applications, Pattern Recognition, 39, 1199-1204, 2005.

Tatlıdil, H., Uygulamalı Çok Değişkenli İstatistik Teknikleri, Cem Ofset Ltd.Şti. Ankara, 1996.

Türker, N., Tokan, F., Yıldırım, T., Kalp Hastalığı Teşhisinde Yapay Sinir Ağlarının Performansının ROC Analizi ile Belirlenmesi, İstanbul National Symposium on Biomedical Engineering, 206-208, 2005.

Ulupınar, S.D., 2001 Kriz Dönemi, Öncesi ve Sonrasında Türk Ticari Bankalarının Karlılıklarının Lojistik Regresyon Analizi ile İncelenmesi, İstatistik Bilim Dalı Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul, 2007.

UNDP, Human Development Report 1990. New York: Oxford University Press,1990.

UNDP, Human Development Report 1991. New York: Oxford University Press,1991.

UNDP, Human Development Report 1994. New York: Oxford University Press,1994.

UNDP, Human Development Report 1996. New York: Oxford University Press, 1996.

UNDP, Web Sitesi http://hdr.undp.org/en/statistics/data/hdi2008. Erişim Tarihi:17.05.2009.

Ülgener, S., Milli Gelir, İstihdam ve İktisadi Büyüme. İstanbul: İstanbul Üniversitesi İşletme Fakültesi Yayınları, 1974.

Ünal, M., Ayırma Analizi ve Bir Uygulama, Yüksek Lisans Tezi, Gazi Üniv. Fen Bilimleri Enstitüsü, 2006.

Veelenturf, L.P.J., Analysis and Application of Artificial Neural Networks, Prentice Hall, 1995.

Warren, S.S., Neural Networks and Statistical Models, Proceeding of the Nineteenth Annual SAS Users Group Intertnational Conference, 1994.

Wu, D.D., Liang, L., Zjiang, Y., Analyzing Financial Distress of Chinese Public Companies Using Probabilistic Neural Networks and Multivariate Discriminate Analysis, Socio Economic Planning Sciences, 42, 206-220, 2008.

Xie, J. ve Qiu, Z., The Effect of Imbalanced Data Sets on LDA:A Theoretical and Empirical Analysis, Pattern Recognition, 40, 557-562, 2006.

Xu, Y., Yang, J., Jin, Z., Theory Analysis on FSLDA and ULDA, The Journal of The Pattern Recognition Society, 36, 3031-3033, 2003.

Xu, Y., Yang, J., Jin, Z., A Novel Method for Fisher Discriminant Analysis, The Journal of The Pattern Recognition Society, 37, 381-384, 2003.

Zheng, W., A Note on Kernel Uncorrelated Discriminant Analysis, The Journal of The Pattern Recognition Society, 38, 2185-2187, 2005.

Zheng, W., Lai, J.H., Li, S.Z., 1D-LDA vs. 2D-LDA:When is Vector-Based Linear Discriminant Analysis Better than Matrix-Based?, Pattern Recognition, 41, 2156-2172, 2007.

ÖZGEÇMİŞ

17.03.1976 yılında Erzurum'da doğdu. İnönü İlkokulunda ilköğrenimine başlayarak, Atatürk İlkokulu'ndan mezun oldu. Orta öğrenimine Gazi Ahmet Muhtar Paşa Ortaokulunda başladı ve Sabancı Ortaokulunda tamamladı.

1989 yılında Kuleli Askeri Lisesi sınavlarını kazanarak lise öğrenimine başladı. 1993 yılında Kara Harp Okulunda başladığı Sistem Mühendisliği lisans programını 1997 yılında başarıyla tamamlayarak meslek hayatına subay naspedilerek atıldı.

Genel Kurmay Başkanlığına bağlı birliklerde başlayan görev safahatı içerisinde KHO Savunma Bilimleri Enstitüsünde Malzeme Tedarik ve Lojistik Yönetimi ABD, Malzeme Tedarik Yönetimi Bilim Dalında tezli yüksek lisansını 2004-2006 yılları arasında tamamladı.

2006 yılında Atatürk Üniversitesi Sosyal Bilimler Enstitüsü İşletme Ana Bilim Dalı Sayısal Yöntemler Bilim Dalında başladığı Doktora öğrenimine halen tez aşamasında devam etmektedir.

2008 yılı atamaları ile Kara Harp Okuluna öğretim elemanı olarak atanmış ve halen Üretim Yönetimi, Karar Analizi ve İstatistik derslerini vermektedir.

Kullanılan Değişkenler:		Toplam Nüfus	Kentleşme Oranı	Sağlık Harcamaları			Doğumda Yaşam Beklentisi	İlkokula Net Kayıt Oranı	1000 kişiye düşen Telefon Hattı Sayısı	1000 Kişiye Düşen Cep Telefonu Aboneliği Sayısı	1000 Kişiye Düşen İnternet Kullanıcı Sayısı	GSYİH	İthal Edilen Mallar ve Hizmetler	İhraç Edilen Mallar ve Hizmetler	Kişi Başına Elektrik Tüketimi kw-h	Hapiste Bulunan Kişi Sayısı	Kadın Parlamenter Oranı
İNSANİ KALKINMIŞLIK SIRALAMASI	Ülkeler/Değişkenlerin elde Edildiği Yıl	(Milyon) 2005	(Toplam Nüfusun Yüzdesi) 2005	(Kamu) GSYİH'nın %'si olarak 2004	(Özel) GSYİH'nın %'si olarak 2004	Kişi Başına Satın Alma Gücü Paritesine Göre 2004	(Yıl) 2000-2005	(%) 2005	2005	2005	2005	Milyar Dolar 2005	GSYİH'nın %'si olarak 2005	GSYİH'nın %'si olarak 2005	2004	(Toplam) 2007	(Toplamın %'si) (2005)
ÇOK GELİŞMİŞ ÜLKELER																	
1	İlanda	0,30	92,80	8,30	1,60	3,294	81,00	99	653	1,024	869	15,80	45	32	29,430	119	31,70
2	Norveç	4,60	77,40	8,10	1,60	4,080	79,30	98	460	1,028	735	295,50	28	45	26,657	3048	37,90
3	Avustralya	20,30	88,20	6,50	3,10	3,123	80,40	97	564	906	698	732,50	21	18	11,849	25353	28,30
4	Kanada	32,30	80,10	6,80	3,00	3,173	79,80	99	566	514	520	1,113,8	34	39	18,408	34096	24,30
5	İrlanda	4,10	60,50	5,70	1,50	2,618	77,80	96	489	1,012	276	201,80	68	83	6,751	3080	14,20
6	İsveç	9,00	84,20	7,70	1,40	2,828	80,10	96	717	935	764	357,70	41	49	16,670	7450	47,30
7	İsviçre	7,40	75,20	6,70	4,80	4,011	80,70	93	689	921	498	367,00	39	46	8,669	6111	24,80
8	Japonya	127,90	65,80	6,30	1,50	2,293	81,90	100	460	742	668	4,534,0	11	13	8,459	79055	11,10
9	Hollanda	16,30	80,20	5,70	3,50	3,092	78,70	99	466	970	739	624,20	63	71	7,196	21013	36,00
10	Fransa	61,00	76,70	8,20	2,30	3,040	79,60	99	586	789	430	2,126,6	27	26	8,231	52009	13,90
11	Finlandiya	5,20	61,10	5,70	1,70	2,203	78,40	98	404	997	534	193,20	35	39	17,374	3954	42,00
12	Amerika	299,80	80,80	6,90	8,50	6,096	77,40	92	606	680	630	12,416,5	15	10	14,240	2186230	16,30
13	İspanya	43,40	76,70	5,70	2,40	2,099	80,00	99	422	952	348	1,124,6	31	25	6,412	64215	30,50
14	Danimarka	5,40	85,60	7,10	1,50	2,780	77,30	95	619	1,010	527	258,70	44	49	6,967	4198	36,90
15	Avusturya	8,30	66,00	7,80	2,50	3,418	78,90	97	450	991	486	306,10	48	53	8,256	8766	31,00
16	İngiltere	60,20	89,70	7,00	1,10	2,560	78,50	99	528	1,088	473	2,198,8	30	26	6,756	88458	19,30
17	Belçika	10,40	97,20	6,90	2,80	3,133	78,20	99	461	903	458	370,80	85	87	8,986	9597	35,70
18	Lüksemburg	0,50	82,80	7,20	0,80	5,178	78,20	95	535	1,576	690	36,50	136	158	16,630	768	23,30
19	Yeni Zelanda	4,10	86,20	6,50	1,90	2,081	79,20	99	422	861	672	109,30	30	29	10,238	7620	32,20
20	İtalya	58,60	67,60	6,50	2,20	2,414	79,90	99	427	1,232	478	1,762,5	26	26	6,029	61721	16,10
21	Çin	7,10	100,00	,,	,,	,,	81,50	93	546	1,252	508	177,70	185	198	6,401	11580	,,
22	Almanya	82,70	75,20	8,20	2,40	3,171	78,70	96	667	960	455	2,794,9	35	40	7,442	78581	30,60
23	İsrail	6,70	91,60	6,10	2,60	1,972	79,70	97	424	1,120	470	123,40	51	46	6,924	13909	14,20
24	Yunanistan	11,10	59,00	4,20	3,70	2,179	78,30	99	568	904	180	225,20	28	21	5,630	9984	13,00
25	Singapur	4,30	100,00	1,30	2,40	1,118	78,80	,,	425	1,010	571	116,80	213	243	8,685	15038	24,50
26	Kore	47,90	80,80	2,90	2,70	1,135	77,00	99	492	794	684	787,60	40	42	7,710	45882	13,40
27	Slovenya	2,00	51,00	6,60	2,10	1,815	76,80	98	408	879	545	34,40	65	65	7,262	1301	10,80
28	Kıbrıs	0,80	69,30	2,60	3,20	1,128	79,00	99	554	949	430	15,40	,,	,,	5,718	580	14,30
29	Portekiz	10,50	57,60	7,00	2,80	1,897	77,20	98	401	1,085	279	183,30	37	29	4,925	12870	21,30
30	Bruney	0,40	73,50	2,60	0,60	621	76,30	93	224	623	277	6,40	,,	,,	8,842	529	,,
31	Barbados	0,30	52,70	4,50	2,60	1,151	76,00	98	500	765	594	3,10	69	58	3,304	997	17,60
32	Çekoslavakya	10,20	73,50	6,50	0,80	1,412	75,40	92	314	1,151	269	124,40	70	72	6,720	18950	15,30

33	Kuveyt	2,70	98,30	2,20	0,60	538	76,90	87	201	939	276	80,80	30	68	15,423	3500	3,10
34	Malta	0,40	95,30	7,00	2,20	1,733	78,60	86	501	803	315	5,60	82	71	5,542	352	9,20
35	Katar	0,80	95,40	1,80	0,60	688	74,30	96	253	882	269	42,50	33	68	19,840	465	0,00
36	Macaristan	10,10	66,30	5,70	2,20	1,308	72,40	89	333	924	297	109,20	69	66	4,070	15720	10,40
37	Polonya	38,20	62,10	4,30	1,90	814	74,60	96	309	764	262	303,20	37	37	3,793	87901	19,10
38	Arjantin	38,70	90,10	4,30	5,30	1,274	74,30	99	227	570	177	183,20	19	25	2,714	54472	36,80
39	Birleşik Arap Emirlikleri	4,10	76,70	2,00	0,90	503	77,80	71	273	1,000	308	129,70	76	94	12,000	8927	22,50
40	Şili	16,30	87,60	2,90	3,20	720	77,90	90	211	649	172	115,20	34	42	3,347	39916	12,70
41	Bahreyn	0,70	96,50	2,70	1,30	871	74,80	97	270	1,030	213	12,90	64	82	11,932	701	13,80
42	Slovakya	5,40	56,20	5,30	1,90	1,061	73,80	92	222	843	464	46,40	83	79	5,335	8493	19,30
43	Litvanya	3,40	66,60	4,90	1,60	843	72,10	89	235	1,275	358	25,60	65	58	3,505	8124	24,80
44	Estonya	1,30	69,10	4,00	1,30	752	70,90	95	328	1,074	513	13,10	90	84	6,168	4463	21,80
45	Latviya	2,30	67,80	4,00	3,10	852	71,30	88	318	814	448	15,80	62	48	2,923	6676	19,00
46	Uruguay	3,30	92,00	3,60	4,60	784	75,30	93	290	333	193	16,80	28	30	2,408	6947	10,80
47	Hırvatistan	4,60	56,50	6,10	1,50	917	74,90	87	425	672	327	38,50	56	47	3,818	3594	21,70
48	Kosta Rika	4,30	61,70	5,10	1,50	592	78,10	,,	321	254	254	20,00	54	48	1,876	7782	38,60
49	Bahama	0,30	90,40	3,40	3,40	1,349	71,10	91	439	584	319	5,50	,,	,,	6,964	1500	22,20
50	Seychelles	0,10	52,90	4,60	1,50	634	,,	99	253	675	249	0,70	121	110	2,716	193	23,50
51	Küba	11,30	75,50	5,50	0,80	229	77,20	97	75	12	17	,,	,,	,,	1,380	55000	36,00
52	Meksika	104,30	76,00	3,00	3,50	655	74,90	98	189	460	181	768,40	32	30	2,130	214450	21,50
53	Bulgaristan	7,70	70,00	4,60	3,40	671	72,40	93	321	807	206	26,60	77	61	4,582	11436	22,10
54	Saint Kitts and Nevis	(,)	32,20	3,30	1,90	710	,,	93	532	213	,,	0,50	61	49	3,333	214	0,00
55	Tonga	0,10	24,00	5,00	1,30	316	72,30	95	,,	161	29	0,20	44	10	327	128	3,30
56	Libya	5,90	84,80	2,80	1,00	328	72,70	,,	133	41	36	38,80	36	48	3,147	11790	7,70
57	Antigua and Barbuda	0,10	39,10	3,40	1,40	516	,,	,,	467	663	350	0,90	69	62	1,346	176	13,90
58	Umman	2,50	71,50	2,40	0,60	419	74,20	76	103	519	111	24,30	43	57	5,079	2020	7,80
59	Trinidad ve Tobago	1,30	12,20	1,40	2,10	523	69,00	90	248	613	123	14,40	46	58	4,921	3851	25,40
60	Romanya	21,60	53,70	3,40	1,70	433	71,30	93	203	617	208	98,60	43	33	2,548	35429	10,70
61	Suudi Arabistan	23,60	81,00	2,50	0,80	601	71,60	78	164	575	70	309,80	26	61	6,902	28612	0,00
62	Panama	3,20	70,80	5,20	2,50	632	74,70	98	136	418	64	15,50	72	69	1,807	11649	16,70
63	Malezya	25,70	67,30	2,20	1,60	402	73,00	95	172	771	435	130,30	100	123	3,196	35644	13,10
64	Belarus	9,80	72,20	4,60	1,60	427	68,40	89	336	419	347	29,60	60	61	3,508	41583	29,80
65	Maritius	1,20	42,40	2,40	1,90	516	72,00	95	289	574	146	6,30	61	57	1,775	2464	17,10
66	Bosna	3,90	45,70	4,10	4,20	603	74,10	,,	248	408	206	9,90	81	36	2,690	1526	14,00
67	Rusya	144,00	73,00	3,70	2,30	583	64,80	92	280	838	152	763,70	22	35	6,425	869814	8,00
68	Arnavutluk	3,20	45,40	3,00	3,70	339	75,70	94	88	405	60	8,40	46	22	1,847	3491	7,10
69	Makedonya	2,00	68,90	5,70	2,30	471	73,40	92	262	620	79	5,80	62	45	3,863	2026	28,30
70	Brezilya	186,80	84,20	4,80	4,00	1,520	71,00	95	230	462	195	796,10	12	17	2,340	361402	9,30
ORTA DÜZEYDE GELİŞMİŞ ÜLKELER																	
71	Dominik	0,10	72,90	4,20	1,70	309	,,	84	293	585	361	0,30	69	45	1,129	289	12,90
72	Saint Lucia	0,20	27,60	3,30	1,80	302	72,50	97	,,	573	339	0,80	70	60	1,879	503	10,30
73	Kazakistan	15,20	57,30	2,30	1,50	264	64,90	91	167	327	27	57,10	45	54	4,320	49292	8,60
74	Venezuella	26,70	93,40	2,00	2,70	285	72,80	91	136	470	125	140,20	21	41	3,770	19853	18,60
75	Kolombiya	44,90	72,70	6,70	1,10	570	71,70	87	168	479	104	122,30	21	21	1,074	62216	9,70
76	Ukrayna	46,90	67,80	3,70	2,80	427	67,60	83	256	366	97	82,90	53	54	3,727	165716	8,70
77	Samoa	0,20	22,40	4,10	1,20	218	70,00	90	73	130	32	0,40	51	27	619	223	6,10
78	Tayland	63,00	32,30	2,30	1,20	293	68,60	88	110	430	110	176,60	75	74	2,020	164443	8,70
79	Dominik Cum.	9,50	66,80	1,90	4,10	377	70,80	88	101	407	169	29,50	38	34	1,536	12725	17,10

80	Belize	0,30	48,30	2,70	2,40	339	75,60	94	114	319	130	1,10	63	55	686	1359	11,90
81	Çin	1,313,0	40,40	1,80	2,90	277	72,00	,,	269	302	85	2,234,3	32	37	1,684	1548498	20,30
82	Grenada	0,10	30,60	5,00	1,90	480	67,70	84	309	410	182	0,50	76	43	1,963	237	28,60
83	Ermenistan	3,00	64,10	1,40	4,00	226	71,40	79	192	106	53	4,90	40	27	1,744	2879	9,20
84	Turkiye	73,00	67,30	5,20	2,10	557	70,80	89	263	605	222	362,50	34	27	2,122	65458	4,40
85	Surinam	0,50	73,90	3,60	4,20	376	69,10	94	180	518	71	1,30	60	41	3,437	1600	25,50
86	Ürdün	5,50	82,30	4,70	5,10	502	71,30	89	119	304	118	12,70	93	52	1,738	5589	7,90
87	Peru	27,30	72,60	1,90	2,20	235	69,90	96	80	200	164	79,40	19	25	927	35642	29,20
88	Lübnan	4,00	86,60	3,20	8,40	817	71,00	92	277	277	196	21,90	44	19	2,691	5971	4,70
89	Ekvador	13,10	62,80	2,20	3,30	261	74,20	98	129	472	47	36,50	32	31	1,092	12251	25,00
90	Filipinler	84,60	62,70	1,40	2,00	203	70,30	94	41	419	54	99,00	52	47	677	89639	22,10
91	Tunus	10,10	65,30	2,80	2,80	502	73,00	97	125	566	95	28,70	51	48	1,313	26000	19,30
92	Fiji	0,80	50,80	2,90	1,70	284	67,80	96	122	229	77	2,70	,,	74	926	1113	,,
93	Saint Vincent and the Grenadines	0,10	45,90	3,90	2,20	418	70,60	90	189	593	84	0,40	65	44	1,030	367	18,20
94	Iran	69,40	66,90	3,20	3,40	604	69,50	95	278	106	103	189,80	30	39	2,460	147926	4,10
95	Paraguay	5,90	58,50	2,60	5,10	327	70,80	88	54	320	34	7,30	54	47	1,146	5063	9,60
96	Gürcistan	4,50	52,20	1,50	3,80	171	70,50	93	151	326	39	6,40	54	42	1,577	11731	9,40
97	Guyana	0,70	28,20	4,40	0,90	329	63,60	,,	147	375	213	0,80	124	88	1,090	1524	29,00
98	Azerbaycan	8,40	51,50	0,90	2,70	138	66,80	85	130	267	81	12,60	54	57	2,796	18259	11,30
99	Sri Lanka	19,10	15,10	2,00	2,30	163	70,80	97	63	171	14	23,50	46	34	420	23613	4,90
100	Maldivler	0,30	29,60	6,30	1,40	494	65,60	79	98	466	59	0,80	110	62	539	1125	12,00
101	Jamaika	2,70	53,10	2,80	2,40	223	72,00	90	129	1,017	404	9,60	61	41	2,697	4913	13,60
102	Cape Verde	0,50	57,30	3,90	1,30	225	70,20	90	141	161	49	1,00	66	32	529	755	15,30
103	El Salvador	6,70	59,80	3,50	4,40	375	70,70	93	141	350	93	17,00	45	27	732	12176	16,70
104	Cezayir	32,90	63,30	2,60	1,00	167	71,00	97	78	416	58	102,30	23	48	889	42000	6,20
105	Viet Nam	85,00	26,40	1,50	4,00	184	73,00	88	191	115	129	52,40	75	70	560	88414	25,80
106	Filistin	3,80	71,60	7,80	5,20	,,	72,40	80	96	302	67	4,00	68	14	513	,,	,,
107	Endonezya	226,10	48,10	1,00	1,80	118	68,60	96	58	213	73	287,20	29	34	476	99946	11,30
108	Suriye	18,90	50,60	2,20	2,50	109	73,10	95	152	155	58	26,30	40	37	1,784	10599	12,00
109	Turkmenistan	4,80	46,20	3,30	1,50	245	62,40	,,	80	11	8	8,10	48	65	2,060	22000	16,00
110	Nikaragua	5,50	59,00	3,90	4,30	231	70,80	87	43	217	27	4,90	58	28	525	5610	18,50
111	Moldovya	3,90	46,70	4,20	3,20	138	67,90	86	221	259	96	2,90	91	53	1,554	8876	21,80
112	Mısır	72,80	42,80	2,20	3,70	258	69,80	94	140	184	68	89,40	33	30	1,465	61845	3,80
113	Özbekistan	26,60	36,70	2,40	2,70	160	66,50	,,	67	28	34	14,00	30	40	1,944	48000	16,40
114	Moğolistan	2,60	56,70	4,00	2,00	141	65,00	84	61	218	105	1,90	84	76	1,260	6998	6,60
115	Honduras	6,80	46,50	4,00	3,20	197	68,60	91	69	178	36	8,30	61	41	730	11589	23,40
116	Kırgızistan	5,20	35,80	2,30	3,30	102	65,30	87	85	105	54	2,40	58	39	2,320	15744	0,00
117	Bolivya	9,20	64,20	4,10	2,70	186	63,90	95	70	264	52	9,30	33	36	493	7710	14,60
118	Guatemala	12,70	47,20	2,30	3,40	256	69,00	94	99	358	79	31,70	30	16	532	7227	8,20
119	Gabon	1,30	83,60	3,10	1,40	264	56,80	77	28	470	48	8,10	39	59	1,128	2750	13,70
120	Vanuatu	0,20	23,50	3,10	1,00	123	68,40	94	33	60	38	,,	,,	,,	206	138	3,80
121	G Afrika	47,90	59,30	3,50	5,10	748	53,40	87	101	724	109	239,50	29	27	4,818	157402	32,80
122	ürdün	6,60	24,70	1,00	3,40	54	65,90	97	39	41	1	2,30	73	54	2,638	10804	19,60
123	Sao Tome and Principe	0,20	58,00	9,90	1,60	141	64,30	97	46	77	131	0,10	99	40	99	155	7,30
124	Botswana	1,80	57,40	4,00	2,40	504	46,60	85	75	466	34	10,30	35	51	,,	6259	11,10
125	Namibya	2,00	35,10	4,70	2,10	407	51,50	72	64	244	37	6,10	45	46	,,	4814	26,90
126	Fas	30,50	58,70	1,70	3,40	234	69,60	86	44	411	152	51,60	43	36	652	54542	6,40

127	Ekvator Ginesi	0,50	38,90	1,20	0,40	223	49,30	81	20	192	14	3,20	,,	,,	52	,,	18,00
128	Hindistan	1,134,4	28,70	0,90	4,10	91	62,90	89	45	82	55	805,70	24	21	618	332112	9,00
129	Solomon Adaları	0,50	17,00	5,60	0,30	114	62,30	63	16	13	8	0,30	46	48	107	297	0,00
130	Lao People's Democratic Republic	5,70	20,60	0,80	3,10	74	61,90	84	13	108	4	2,90	31	27	126	4020	25,20
131	Kamboçya	14,00	19,70	1,70	5,00	140	56,80	99	3	75	3	6,20	74	65	10	8160	11,40
132	Myanmar	48,00	30,60	0,30	1,90	38	59,90	90	9	4	2	,,	,,	,,	129	60000	,,
133	Bhutan	0,60	11,10	3,00	1,60	93	63,50	,,	51	59	39	0,80	55	27	229	,,	2,70
134	Comoros	0,80	37,00	1,60	1,20	25	63,00	55	28	27	33	0,40	35	12	31	200	3,00
135	Gana	22,50	47,80	2,80	3,90	95	58,50	65	15	129	18	10,70	62	36	289	12736	10,90
136	Pakistan	158,10	34,90	0,40	1,80	48	63,60	68	34	82	67	110,70	20	15	564	89370	20,40
137	Moritanya	3,00	40,40	2,00	0,90	43	62,20	72	13	243	7	1,90	95	36	112	815	17,60
138	Lesotho	2,00	18,70	5,50	1,00	139	44,60	87	27	137	24	1,50	88	48	,,	2924	25,00
139	Kongo	3,60	60,20	1,20	1,30	30	53,00	44	4	123	13	5,10	55	82	229	918	10,10
140	Bangladeş	153,30	25,10	0,90	2,20	64	62,00	94	8	63	3	60,00	23	17	154	71200	15,10
141	Swaziland	1,10	24,10	4,00	2,30	367	43,90	80	31	177	32	2,70	95	88	,,	2734	16,80
142	Nepal	27,10	15,80	1,50	4,10	71	61,30	79	17	9	4	7,40	33	16	86	7135	17,30
143	Madagaskar	18,60	26,80	1,80	1,20	29	57,30	92	4	27	5	5,00	40	26	56	20294	8,40
144	Kamerun	17,80	54,60	1,50	3,70	83	49,90	,,	6	138	15	16,90	25	23	256	20000	8,90
145	Papua Yeni Gine	6,10	13,40	3,00	0,60	147	56,70	,,	11	4	23	4,90	54	45	620	4056	0,90
146	Haiti	9,30	38,80	2,90	4,70	82	58,10	,,	17	48	70	4,30	45	16	61	3670	6,30
147	Sudan	36,90	40,80	1,50	2,60	54	56,40	43	18	50	77	27,50	28	18	116	12000	16,40
148	Kenya	35,60	20,70	1,80	2,30	86	51,00	79	8	135	32	18,70	35	27	169	47036	7,30
149	Djibouti	0,80	86,10	4,40	1,90	87	53,40	33	14	56	13	0,70	54	37	260	384	10,80
150	Timor-Leste	1,10	26,50	8,80	2,40	143	58,30	98	,,	,,	,,	0,30	,,	,,	294	320	25,30
151	Zimbabwe	13,10	35,90	3,50	4,00	139	40,00	82	25	54	77	3,40	53	43	924	18033	22,20
152	Togo	6,20	40,10	1,10	4,40	63	57,60	78	10	72	49	2,20	47	34	102	3200	8,60
153	Yemen	21,10	27,30	1,90	3,10	82	60,30	75	39	95	9	15,10	38	46	208	14000	0,70
154	Uganda	28,90	12,60	2,50	5,10	135	47,80	,,	3	53	17	8,70	27	13	63	26126	29,80
155	Gambiya	1,60	53,90	1,80	5,00	88	58,00	77	29	163	33	0,50	65	45	98	450	9,40